

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1299	100.0	242	8	ADJ93360	Adj93360	Human BGS
2	1299	100.0	541	8	ADJ93365	Adj93365	Human tub
3	1299	100.0	541	8	ADJ93358	Adj93358	Human BGS
4	1247	96.0	293	8	ADJ93366	Adj93366	Human BGS
5	850	65.4	352	3	AAB43005	Aab43005	Human ORF
6	850	65.4	352	4	AAM39450	Aam39450	Human pol
7	850	65.4	352	8	ADJ93457	Adj93457	Human HOT
8	850	65.4	352	8	ABM80420	Abm80420	Tumour-as
9	850	65.4	362	3	AAB58909	Aab58909	Breast an
10	850	65.4	399	6	ABU11512	Abu11512	Human MDD
11	850	65.4	488	5	AAB74334	Aau74334	Human cyt
12	850	65.4	744	4	AAB94796	Aab94796	Human pro
13	847	65.2	326	7	ADM05524	Adm05524	Human pro
14	846	65.1	282	8	ADJ93455	Adj93455	Human HOT
15	628.5	48.4	992	4	ABB56645	Abb56645	Drosophil
16	555	42.7	432	6	ABU92048	Abu92048	Human pro
17	499.5	38.5	719	4	ABB65541	Abb65541	Drosophil
18	468	36.0	281	8	ADQ66614	Adq66614	Novel hum
19	393	30.3	362	4	AAM41236	Aam41236	Human pol
20	337	25.9	566	4	ABG05971	Abg05971	Novel hum
21	324.5	25.0	347	6	ABU00150	Abu00150	Human nov
22	324	24.9	496	4	ABB64074	Abb64074	Drosophil
23	297.5	22.9	827	4	ABB60840	Abb60840	Drosophil
24	296	22.8	1281	4	AAM39105	Aam39105	Human pol

XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
 CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytostatic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
 CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 CC immunosuppressive, anticancer or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of the human
 CC BGS-42 protein (partial sequence) of the invention.
 XX
 SQ Sequence 242 AA;

Query Match 100.0%; Score 1299; DB 8; Length 242;
 Best Local Similarity 100.0%; Pred. No. 1.9e-134;
 Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 IDGLRNIWIKPAKSRGRDIVCMRVRETELEAAADHPLSRDNKWKVQKIETPLICD 60
 Db 1 IDGLRNIWIKPAKSRGRDIVCMRVRETELEAAADHPLSRDNKWKVQKIETPLICD 60
 Qy 61 TKFDIRQWFLVTDWNPITWIFKESYLRFSTQFSLDKLDSAIHLCCNNAVQKYLKNDVGR 120
 Db 61 TKFDIRQWFLVTDWNPITWIFKESYLRFSTQFSLDKLDSAIHLCCNNAVQKYLKNDVGR 120
 Qy 121 SPILLPAHNMTSTRFOEYLQQRGAVGWSVIYPSMKKATAHAKVQAQDHVEPRKNSFEL 180
 Db 121 SPILLPAHNMTSTRFOEYLQQRGAVGWSVIYPSMKKATAHAKVQAQDHVEPRKNSFEL 180
 Qy 181 YGADFVLGRDFRPLWLEINSPTWHPSTPTVTAQLCAQVQEDTIKVAVDRCSDIGNFELLW 240
 Db 181 YGADFVLGRDFRPLWLEINSPTWHPSTPTVTAQLCAQVQEDTIKVAVDRCSDIGNFELLW 240
 Qy 241 RQ 242
 Db 241 RQ 242

RESULT 2
 ADJ93365
 ID ADJ93365 standard; protein; 541 AA.

XX ADJ93365;

XX 06-MAY-2004 (first entry)

XX Human tubulin tyrosine ligase protein consensus sequence SeqID13.

XX testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytostatic; respiratory-Gen; gastrointestinal-Gen;
 KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; anticancer;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;

KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human.
 XX
 OS Homo sapiens.
 OS Synthetic.

XX WO2004005487-A2.

XX 15-JAN-2004.

XX 09-JUL-2003; 2003WO-US021605.

XX 09-JUL-2002; 2002US-0394725P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Feder JN, Wu S, Nelson TC;

XX WPI; 2004-099381/10.

XX N-PSDB; ADJ93364.

XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
 PT useful for preventing, treating or ameliorating a medical condition, e.g.
 PT aberrant cellular proliferation, reproductive disorders or testicular
 PT disorders.
 XX

XX Example 4; SEQ ID NO 13; 343pp; English.

XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
 CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytostatic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
 CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 CC immunosuppressive, anticancer or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of the
 CC tubulin tyrosine ligase protein consensus sequence which was used in the
 CC exemplification of the invention.
 XX

XX Sequence 541 AA;

Query Match 100.0%; Score 1299; DB 8; Length 541;
 Best Local Similarity 100.0%; Pred. No. 6e-134;
 Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IDGLRNIWIKPAKSRGRDIVCMRVRETELEAAADHPLSRDNKWKVQKIETPLICD 60

Db 133 IDGLRNIWIKPAKSRGRDIVCMRVRETELEAAADHPLSRDNKWKVQKIETPLICD 192

Qy 61 TKFDIRQWFLVTDWNPITWIFKESYLRFSTQFSLDKLDSAIHLCCNNAVQKYLKNDVGR 120

Db 193 TKFDIRQWFLVTDWNPITWIFKESYLRFSTQFSLDKLDSAIHLCCNNAVQKYLKNDVGR 252

QY 121 SPLLPAHNNMTSTRFOEYLQQRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180
 DB 253 SPLLPAHNNMTSTRFOEYLQQRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
 QY 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKVAVDRSCDIGNFELW 240
 DB 313 YGADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKVAVDRSCDIGNFELW 372
 QY 241 RQ 242
 DB 373 RQ 374

RESULT 3
 ID ADJ93358
 XX ADJ93358 standard; protein; 541 AA.
 AC ADJ93358;
 XX
 DT 06-MAY-2004 (first entry)
 DE Human BGS-42 protein sequence SeqID2.
 DE
 XX testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
 KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
 KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human.
 XX Homo sapiens.
 OS
 PN WO2004005487-A2.
 XX
 PD 15-JAN-2004.
 XX
 PF 09-JUL-2003; 2003WO-US021605.
 XX
 PR 09-JUL-2002; 2002US-0394725P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Feder JN, Wu S, Nelson TC;
 XX
 DR WPI, 2004-099381/10.
 DR N-FSDB; ADJ93357.
 XX
 XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
 PT useful for preventing, treating or ameliorating a medical condition, e.g.
 PT aberrant cellular proliferation, reproductive disorders or testicular
 PT disorders.
 XX
 FS Claim 5; SEQ ID NO 2; 343pp; English.
 XX
 CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
 CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytosolic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
 CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 CC immunosuppressive, antiseborrheic or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC used for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a

CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of the human
 CC BGS-42 protein of the invention.
 XX
 SQ Sequence 541 AA;

Query Match 100.0%; Score 1299; DB 8; Length 541;
 Best Local Similarity 100.0%; Pred. No. 6e-134;
 Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IDGLNIIWIIPAAKSRGRDIVCMRVBEILELAADHPLSRDNKVVQKYIETPLLICD 60
 DB 133 IDGLNIIWIIPAAKSRGRDIVCMRVBEILELAADHPLSRDNKVVQKYIETPLLICD 192
 QY 61 TKFDIRQWFLVTDNPLTIWFYKESYLRFTQRFSLDKLDSAIHLICNNAVQKYLKNDVGR 120
 DB 193 TKFDIRQWFLVTDNPLTIWFYKESYLRFTQRFSLDKLDSAIHLICNNAVQKYLKNDVGR 252
 QY 121 SPLPAHNNMTSTRFOEYLQQRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180
 DB 253 SPLPAHNNMTSTRFOEYLQQRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
 QY 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKVAVDRSCDIGNFELW 240
 DB 313 YGADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKVAVDRSCDIGNFELW 372
 QY 241 RQ 242
 DB 373 RQ 374

RESULT 4
 ID ADJ93366 standard; protein; 293 AA.
 AC ADJ93366;
 DT 06-MAY-2004 (first entry)
 DE Human BGS-42 protein-related TTL1 domain.
 DE
 XX

testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
 KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
 KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human; TTL1 domain.
 XX Homo sapiens.

WO2004005487-A2.

QY 1 IDGLRNIWIKPAKSRGRDIVCMRVEIELEAAADHPLSRDNKWWVQKIETPLLCD 60
 DB 61 MEGDRNIWIKPAKSRGRGIMCMDEMLKLVNGNVPVMDKGKWWVQKIETPLIFG 120
 QY 61 TKFDIRQWFLVTDNPLTIWFYKESYLRFSTQSFSLDKLDSAIHLCLNNAVQKYLKNDVGR 120
 DB 121 TKFDLRQWFLVTDNPLTIWFYKESYLRFSTQSFSLDKLDSVHLCLNNSIQHLENSCHR 180
 QY 121 SPLPANHMMWTSTFQEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVPRKNSFEL 180
 DB 181 HPLLPDNNWSSQRFQHLQEMGAPNAWSTIIVPGMKDAVIHALQTSQDTVQCRKASFEL 240
 QY 181 YGADFLVGRDPRFWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKAV----DRSCDIGNF 236
 DB 241 YGADFVFGEDFQFWLIEINASPTMAPSTAVTARLCAGVQADTLRVVIDRLDRNCDTGAF 300
 QY 237 ELLMRQ 242
 DB 301 ELIYKQ 306

RESULT 6
 AAM39450
 ID AAM39450 standard; protein; 352 AA.
 AC AAM39450;
 XX
 XX
 DT 22-OCT-2001 (first entry)
 DE Human polypeptide SEQ ID NO 2595.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US034263.
 XX
 PR 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-00552317.
 PR 20-JUN-2000; 2000US-00598042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
 PI Zhou P, Goodrich R, Drmanac RT;
 XX WPI; 2001-442253/47.
 DR N-PSDB; AAI58606.
 XX
 XX Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.
 XX
 PS Example 4; SEQ ID NO 2595; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
 CC encoded polypeptides (AAM38642-AAM42213) with nootropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification

XX Sequence 352 AA;

Query Match 65.4%; Score 850; DB 4; Length 352;
 Best Local Similarity 61.8%; Pred. No. 1.2e-84;
 Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1;

QY 1 IDGLRNIWIKPAKSRGRDIVCMRVEIELEAAADHPLSRDNKWWVQKIETPLLCD 60
 DB 61 MEGDRNIWIKPAKSRGRGIMCMDEMLKLVNGNVPVMDKGKWWVQKIETPLIFG 120
 QY 61 TKFDIRQWFLVTDNPLTIWFYKESYLRFSTQSFSLDKLDSAIHLCLNNAVQKYLKNDVGR 120
 DB 121 TKFDLRQWFLVTDNPLTIWFYKESYLRFSTQSFSLDKLDSVHLCLNNSIQHLENSCHR 180
 QY 121 SPLPANHMMWTSTFQEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVPRKNSFEL 180
 DB 181 HPLLPDNNWSSQRFQHLQEMGAPNAWSTIIVPGMKDAVIHALQTSQDTVQCRKASFEL 240
 QY 181 YGADFLVGRDPRFWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKAV----DRSCDIGNF 236
 DB 241 YGADFVFGEDFQFWLIEINASPTMAPSTAVTARLCAGVQADTLRVVIDRLDRNCDTGAF 300
 QY 237 ELLMRQ 242
 DB 301 ELIYKQ 306

RESULT 7
 ADJ93457
 ID ADJ93457 standard; protein; 352 AA.
 AC ADJ93457;
 XX
 XX
 DT 06-MAY-2004 (first entry)
 DE Human HOTT13 protein sequence SeqID2.

XX testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytostatic; respiratory-gen; gastrointestinal-Gen;
 KW neuroprotective; endocrine-gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
 KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human; HOTT13.

XX Homo sapiens.
 OS
 PN WO2004005487-A2.
 XX
 XX 15-JAN-2004.
 PD
 XX 09-JUL-2003; 2003WO-US021605.
 PF

```
XX 09-JUL-2002; 2002US-0394725P.
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX Feder JN, Wu S, Nelson TC;
XX WPI; 2004-099381/10.
XX
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
XX useful for preventing, treating or ameliorating a medical condition, e.g.
XX aberrant cellular proliferation, reproductive disorders or testicular
XX disorders.
XX
XX Disclosure; SEQ ID NO 7; 343pp; English.
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
XX -like polypeptide, designated the BGS-42 polypeptide. The invention may
XX be useful for the development of compounds with a cytostatic, respiratory
XX -gen, gastrointestinal-gen, neuroprotective, endocrine-gen,
XX antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
XX antiparkinsonian, antithrombotic, antiasthmatic, anti-HIV, antibacterial,
XX immunosuppressive, antiseborrheic or dermatological activity acting as
XX tyrosine ligase modulators. In addition, the disclosed sequences may be
XX useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
XX used for diagnosing a pathological condition or a susceptibility to a
XX pathological condition in a subject, and for preventing, treating or
XX ameliorating a medical condition, such as a disorder related to aberrant
XX tubulin ligase activity, a disorder related to aberrant tubulin-
XX carboxypeptidase activity, aberrant cellular proliferation, reproductive
XX disorders, testicular disorders, testicular cancer, pulmonary disorders,
XX lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
XX neural disorders, brain cancer, liver cancer, or proliferative condition
XX of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
XX polypeptide, polynucleotide, or their modulators are also useful for
XX treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
XX disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
XX -42 polypeptide can be used as a preventive agent for immunological
XX disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
XX disease or scleroderma. The antibodies may be used to purify, detect and
XX target the BGS-42 polypeptides. The present sequence is that of the human
XX HOTT13 protein which is related to the invention. Note: The present
XX sequence does not appear in the specification but was obtained from
XX Genbank.
XX
XX SQ Sequence 352 AA;
XX
XX Query Match 65.4%; Score 850; DB 8; Length 352;
XX Best Local Similarity 61.8%; Pred. No. 1.2e-84;
XX Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1;
XX
XX QY 1 IDGLRNIIWIKPAKSRGRDVCMDRVEEILELAADHPLSRDNKWKVQKIETPLICD 60
XX Db : : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
XX 61 MEGDRNIWIKVPGAKSRGRGIMCMCHLEMLKLVGNPNVWMDGKWKVQKIERPLLIIFG 120
XX
XX QY 61 TKFDIRQWFLVTDWNPITWIFYKESYLRFSTQRPFLDKLSAHLNNAVQKYLKNDVGR 120
XX Db : : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
XX 121 TKFDLRQWFLVTDWNPITWIFYRDSYIRFSTQFPFLKXNLSVHLNNSIQKHLENSCHR 180
XX
XX QY 121 SPLLPALHNMWTSRFBQYLRQREGAVGWSVIVPSMKKATAHAKVQADHVEPRKNSFEL 180
XX Db : : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
XX 181 HPLLPDNNWSSQRFQLEMGAPNAWSTIIIVPGMKDAVIHALQTSQDTVQCKRASFEL 240
XX
XX QY 181 YGADFVLGRFRPWLIEINSPTMHPSTPTVAQICAOVQEDTIKVAV----DRSCDIGNF 236
XX Db : : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
XX 241 YGADFVGEDEFQFPLIINASPTNAPSTAVTARLCAGVQADTLRWVIDRMLDRNCDTGAF 300
XX
XX QY 237 ELLWRQ 242
XX Db : : | | | |
XX 301 ELIYKQ 306
```

RESULT 8

```
ABM80420
ID ABM80420 standard; protein; 352 AA.
XX
XX AC ABM80420;
XX
XX DT 18-NOV-2004 (first entry)
XX
XX DE Tumour-associated antigenic target (TAT) polypeptide PRO80956, SEQ:1058.
XX
XX KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
XX tumour; diagnosis; cell proliferative disorder; breast cancer;
XX colorectal cancer; lung cancer; ovarian cancer; liver cancer;
XX central nervous system cancer; bladder cancer; pancreatic cancer;
XX cervical cancer; melanoma; leukaemia; hybridisation probe;
XX chromosome identification; chromosome mapping; gene mapping;
XX gene therapy; cytostatic.
XX
XX OS Homo sapiens.
XX
XX PN WO2004030615-A2.
XX
XX PD 15-APR-2004.
XX
XX PF 29-SEP-2003; 2003WO-US028547.
XX
XX PR 02-OCT-2002; 2002US-0414971P.
XX
XX (GETH ) GENENTECH INC.
XX
XX Wu TD, Zhang Z, Zhou Y;
XX
XX WPI; 2004-347921/32.
XX N-PSDB; ACN37881.
XX
XX New tumor-associated antigenic target polypeptides and nucleic acids,
XX useful in preparing a medicament for treating or detecting a
XX proliferative disorder, e.g. breast, lung, colorectal, ovarian or
XX prostate cancer or tumor.
XX
XX Claim 12; SEQ ID NO 1058; 7273pp; English.
XX
XX The invention relates to human tumour-associated antigenic target (TAT)
XX polypeptides, and their related nucleic acids. The TAT polypeptides are
XX overexpressed in cancer tissues compared to normal tissues, and may thus
XX serve as effective targets for the diagnosis and treatment of cancer in
XX mammals. The invention also relates to nucleic acid and polypeptide
XX sequences at least 80% identical to the TAT nucleic acids and
XX polypeptides; expression vectors and host cells comprising a TAT nucleic
XX acid; an antibody specific for a TAT polypeptide; a peptide or organic
XX molecule which binds to a TAT polypeptide; fusion proteins comprising a
XX TAT polypeptide; and methods and compositions for the treatment or
XX diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
XX polypeptides, binding molecules and host cells comprising a TAT nucleic
XX acid for diagnosing or treating a cell proliferative disorder associated with
XX increased TAT expression, particularly cancers such as breast cancer,
XX colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
XX cancer, pancreatic cancer, cervical cancer, cancers of the central
XX nervous system, melanoma and leukaemia. TAT nucleic acids may further be
XX used as hybridisation probes, in chromosome and gene mapping, in
XX chromosome identification and in gene therapy. The present sequence
XX represents a TAT polypeptide of the invention
XX
XX SQ Sequence 352 AA;
XX
XX Query Match 65.4%; Score 850; DB 8; Length 352;
XX Best Local Similarity 61.8%; Pred. No. 1.2e-84;
XX Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1;
XX
XX QY 1 IDGLRNIIWIKPAKSRGRDVCMDRVEEILELAADHPLSRDNKWKVQKIETPLICD 60
XX Db : : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
XX 61 MEGDRNIWIKVPGAKSRGRGIMCMCHLEMLKLVGNPNVWMDGKWKVQKIERPLLIIFG 120
XX
XX QY 61 TKFDIRQWFLVTDWNPITWIFYKESYLRFSTQRPFLDKLSAHLNNAVQKYLKNDVGR 120
```

Db 121 TKFDLRQWFLVTDWNPVTWVYRDSYRSTQPSLKNLDNSVHLNCNNSIQHLENSCHR 180
QY 121 SPLLPAHNMTSTRFOEYLQROGKAVGWSVIYPSMKKAIAHAMKVAQDHVPRKNSFEL 180
Db 181 HPLLPPDNWSSQRFQAHLEMGAPNASTIIVPGMKDAVIHALQTSQDTVQCRKASFEL 240
QY 181 YGADFVLGRDPRFWLIEINSSPTMHPSTPTAQLCAQVQEDTIKVAV-----DRSCDIGNF 236
Db 241 YGADFVFGEDFQWLEINASPTMAPSTAVTARLCAGVQADTLRVVIDRMLDRNCDTGAF 300
QY 237 ELLWRQ 242
Db 301 ELIYKQ 306

RESULT 9

AAB58909
ID AAB58909 standard; protein; 362 AA.
XX AAB58909;
DT 27-MAR-2001 (first entry)
DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 617.
XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW neutropenic; neurprotection; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antiulcer; vulnery; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW cardiovascular disorder; wound healing; neurological disease.

OS Homo sapiens.

XX WO200055173-A1.

PN 21-SEP-2000.

PD 08-MAR-2000; 2000WO-US005881.

PF 12-MAR-1999; 99US-0124270P.

PR (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

PI WPI; 2000-611515/58.

XX N-PSDB; AAF21812.

XX New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention, treatment
PT and diagnosis of cancer, immune disorders, cardiovascular disorders and
PT neurological diseases.

PS Claim 11; Page 1056-1057; 1299pp; English.

XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC isolation and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist
CC or antagonist sequences exhibit cytostatic; immunosuppressive; neutropenic;
CC neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
CC antiinflammatory; antiulcer; vulnery; anticonvulsant; antibacterial;
CC antifungal; antiparasitic and cardiant activity. The polynucleotide and
CC protein sequences are used in the diagnosis of cancer, particularly
CC breast and ovarian cancer. The nucleic acid sequences, proteins, agonists
CC and agonists may also be used in the diagnosis, prevention and treatment
CC of immune disorders e.g. Addison's disease, allergies, autoimmune

CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC cardiovascular disorders such as myocardial ischaemias; wound healing;
CC neurological diseases such as cerebral anoxia and epilepsy; and
XX infectious diseases

SQ Sequence 362 AA;

Query Match 65.4%; Score 850; DB 3; Length 362;
Best Local Similarity 61.8%; Pred. No. 1.3e-84;
Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1;

- QY 1 IDGLRNIIWIKPAKSRGRDIVCMRDVBEILELAADHPLSRDNKWKVQKIETPLLICD 60
Db 78 MEGDRNIWIKPAKSRGRGIMCMBDHEMLKLVGNPVMKDGKWKVQKIETPLIFG 137
QY 61 TKFDLRQWFLVTDWNPVTWVYRDSYRSTQPSLKNLDNSVHLNCNNSIQHLENSCHR 120
Db 138 TKFDLRQWFLVTDWNPVTWVYRDSYRSTQPSLKNLDNSVHLNCNNSIQHLENSCHR 197
QY 121 SPLLPAHNMTSTRFOEYLQROGKAVGWSVIYPSMKKAIAHAMKVAQDHVPRKNSFEL 180
Db 198 HPLLPPDNWSSQRFQAHLEMGAPNASTIIVPGMKDAVIHALQTSQDTVQCRKASFEL 257
QY 181 YGADFVLGRDPRFWLIEINSSPTMHPSTPTAQLCAQVQEDTIKVAV-----DRSCDIGNF 236
Db 258 YGADFVFGEDFQWLEINASPTMAPSTAVTARLCAGVQADTLRVVIDRMLDRNCDTGAF 317
QY 237 ELLWRQ 242
Db 318 ELIYKQ 323

RESULT 10

ABU11512
ID ABU11512 standard; protein; 399 AA.

XX AC ABU11512;

XX DT 12-FEB-2003 (first entry)

XX DE Human MDDT polypeptide SEQ ID 459.

XX MDDT; human; disease detection and treatment molecule polypeptide;
KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;
KW haemostatic; nephrotropic; antianaemic; anipsoriatic; hepatotropic;
KW gene therapy; protein replacement therapy; cell proliferative disorder;
KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; myeloma; sarcoma;
KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;
KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;
KW psoriasis; hepatitis.

OS Homo sapiens.

XX WO200279449-A2.

XX PD 10-OCT-2002.

XX PF 27-MAR-2002; 2002WO-US009944.

XX PR 28-MAR-2001; 2001US-0279619P.

XX PR 29-MAR-2001; 2001US-0280067P.

XX PR 29-MAR-2001; 2001US-0280068P.

XX PR 16-MAY-2001; 2001US-0291280P.

XX PR 17-MAY-2001; 2001US-0291829P.

XX PR 19-JUN-2001; 2001US-0291849P.

XX PR 20-JUN-2001; 2001US-0299776P.

XX PR 20-JUN-2001; 2001US-0300001P.

XX (INCY-) INCYTE GENOMICS INC.

XX DAFFO A, JONES AL, TRAN AB, DAHL CR, GIETZEN D, CHINN J;

QY 181 YGADFVLGRDRPRLWLEINSSPTMHPSTPTAQLCAQVQEDTIKAV-----DRSCDIGNP 236
Db 241 YGADFVFGEDFQPLWLEINASPTMAPSTAVTARLCAGVQADTLRVVVIDRMLDRNCDTGAF 300
QY 237 ELLWRQ 242
Db 301 ELIYKQ 306
RESULT 12
ID AAB94796
XX AAB94796 standard; protein; 744 AA.
AC AAB94796;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:15921.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX
XX Claim 8; SEQ ID NO 15921; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
XX
SQ Sequence 744 AA;

Query Match 65.4%; Score 850; DB 4; Length 744;
Best Local Similarity 61.8%; Pred. No. 3.7e-84;
Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1;
QY 1 IDGLRNIIWKPAKSRGRDIVCMRVEREILSLAADHPLSRDNKVVVKYIETPLLICD 60
Db 273 MEGDRNIWIKPFGAKSRGIRGIMDHLEMLKLVNNGPVVMKDGKVVVKYIERPLIFG 332
QY 61 TKFDIRQWFLVTDNPLTIWFYKESYLRSTQRFSLDKLSDSAIHLCCNNAVQKYLKNDVGR 120
Db 333 TKFDLRQWFLVTDNPLTIWFYRDSYIRFSTQPFSLKNLSDNVHLCCNNSIQHLENSCHR 392
QY 121 SPLPAHNNWTSRFEQYLQRCRGAVMGSVIYPSMKKAIAHAMKVAQDHVPRKNSFEL 180
Db 393 HELLPPDNWSSORFOAHLQEMGAPNAWSTIIVPGMKDAVIHALQTSQDTVQCRKASFEL 452
QY 181 YGADFVLGRDRPRLWLEINSSPTMHPSTPTAQLCAQVQEDTIKAV-----DRSCDIGNP 236
Db 453 YGADFVFGEDFQPLWLEINASPTMAPSTAVTARLCAGVQADTLRVVVIDRMLDRNCDTGAF 512
QY 237 ELLWRQ 242
Db 513 ELIYKQ 518
RESULT 13
ADM05524
ID ADM05524 standard; protein; 326 AA.
XX
AC ADM05524;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human protein of the invention SEQ ID NO:4209.
XX
KW human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS Homo sapiens.
XX
PN EP1347046-A1.
XX
PD 24-SEP-2003.
XX
PF 12-APR-2002; 2002EP-00008400.
XX
PR 22-MAR-2002; 2002JP-00137785.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX WPI; 2003-723558/69.
DR N-PSDB; ADM03081.
XX
XX New polynucleotides and polypeptides are useful in gene therapy, for developing a diagnostic marker or medicines for regulating their expression and activity, or as a target of gene therapy.
XX
XX Claim 1; SEQ ID NO 4209; 305pp; English.
XX
XX The invention relates to a novel human polynucleotide and the encoded polypeptide. A polynucleotide of the invention may have a use in gene therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful as a primer for synthesizing the polynucleotide or as a probe for detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are useful in gene therapy, for developing a diagnostic marker or medicines for regulating their expression and activity, or as a target of gene therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides are useful as pharmaceutical agents. The present sequence represents a protein sequence of the invention.

```
XX SQ Sequence 326 AA;
Query Match 65.2%; Score 847; DB 7; Length 326;
Best Local Similarity 61.8%; Pred. No. 2.4e-84;
Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1;
QY 1 IDGLRNWIKPAKSRGRDIVCMRVEIELELAADHPLSRDNKVVQKYIETPLLICD 60
Db :|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
61 MEGDRNIWIKPGAKSRGRGIMCMHLEMLKLVNGNPMVKDGKVVQKYIERPLLIIFG 120
QY 61 TKFDIQWFLVTDWNPPLTTWFKYESYLRFSTQFSLKNDNSVHLNCNSIQHLENSCHR 120
Db |||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
121 TKFDLQWFLVTDWNPPLTVWFRDYSIRFSTQFSLKNDNSVHLNCNSIQHLENSCHR 180
QY 121 SPLLPAHNMTWTRFQEYLQQRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180
Db |||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
181 HPLLPDPNMWSSQRFQAHLEMGAPNAWSTIIIVPGMKDAVIHALQTSQDTVQCRKASFEL 240
QY 181 YGADFVLGRDPRFWLIEINSSPTMHFSTPVTALCAQVQEDTIKVAV----DRSCDIGNF 236
Db |||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
241 YGADFVFGEDFQWFLIEINASPTMAPSTAVTARLCAGVQADTLRVVIDRLDRNCDTGAF 300
QY 237 ELLMRQ 242
Db ||:::|
301 ELIYKQ 306

RESULT 14
ADJ93455
ID ADJ93455 standard; protein; 292 AA.
XX AC ADJ93455;
XX DT 06-MAY-2004 (first entry)
XX DE Human HOTTL protein sequence SeqID2.
XX testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytostatic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; neotropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-Hiv; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; HOTTL.
XX OS Homo sapiens.
XX WO2004005487-A2.
XX 15-JAN-2004.
XX 09-JUL-2003; 2003WO-US021605.
XX 09-JUL-2002; 2002US-0394725P.
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX Feder JN, Wu S, Nelson TC;
XX WPI; 2004-099381/10.
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
XX useful for preventing, treating or ameliorating a medical condition, e.g.
XX aberrant cellular proliferation, reproductive disorders or testicular
XX disorders.
```

```
XX Disclosure; SEQ ID NO 5; 343pp; English.
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
XX -like polypeptide, designated the BGS-42 polypeptide. The invention may
XX be useful for the development of compounds with a cytostatic, respiratory
XX -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
XX antiinflammatory, anabolic, hypertensive, osteopathic, neotropic,
XX antiparkinsonian, antiarthritic, antiasthmatic, anti-Hiv, antibacterial,
XX immunosuppressive, antiseborrheic or dermatological activity acting as
XX tyrosine ligase modulators. In addition, the disclosed sequences may be
XX useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
XX used for diagnosing a pathological condition or a susceptibility to a
XX pathological condition in a subject, and for preventing, treating or
XX ameliorating a medical condition, such as a disorder related to aberrant
XX tubulin ligase activity, a disorder related to aberrant tubulin-
XX carboxypeptidase activity, aberrant cellular proliferation, reproductive
XX disorders, testicular disorders, testicular cancer, pulmonary disorders,
XX lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
XX neural disorders, brain cancer, liver cancer, or proliferative condition
XX of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
XX polypeptide, polynucleotide, or their modulators are also useful for
XX treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
XX disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
XX -42 polypeptide can be used as a preventive agent for immunological
XX disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
XX disease or scleroderma. The antibodies may be used to purify, detect and
XX target the BGS-42 polypeptides. The present sequence is that of the human
XX HOTTL protein which is related to the invention. Note: The present
XX sequence does not appear in the specification but was obtained from
XX Genbank.
XX SQ Sequence 292 AA;
Query Match 65.1%; Score 846; DB 8; Length 292;
Best Local Similarity 61.4%; Pred. No. 2.6e-84;
Matches 151; Conservative 41; Mismatches 50; Indels 4; Gaps 1;
QY 1 IDGLRNWIKPAKSRGRDIVCMRVEIELELAADHPLSRDNKVVQKYIETPLLICD 60
Db :|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
1 MEGDRNIWIKPGAKSRGRGIMCMHLEMLKLVNGNPMVKDGKVVQKYIERPLLIIFG 60
QY 61 TKFDIQWFLVTDWNPPLTTWFKYESYLRFSTQFSLKNDNSVHLNCNSIQHLENSCHR 120
Db |||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
61 TKFDLQWFLVTDWNPPLTVWFRDYSIRFSTQFSLKNDNSVHLNCNSIQHLENSCHR 120
QY 121 SPLLPAHNMTWTRFQEYLQQRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180
Db |||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
181 HPLLPDPNMWSSQRFQAHLEMGAPNAWSTIIIVPGMKDAVIHALQTSQDTVQCRKASFEL 180
QY 181 YGADFVLGRDPRFWLIEINSSPTMHFSTPVTALCAQVQEDTIKVAV----DRSCDIGNF 236
Db |||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
181 YGADFVFGEDFQWFLIEINASPTMAPSTAVTARLCAGVQADTLRVVIDRLDRNCDTGAF 240
QY 237 ELLMRQ 242
Db ||:::|
241 ELIYKQ 246

RESULT 15
ABB65645
ID ABB65645 standard; protein; 992 AA.
XX AC ABB65645;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 23727.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
```

XX WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL09748.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX PT interactions.
XX PS Disclosure; SEQ ID NO 23727; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 992 AA;
Query Match 48.4%; Score 628.5; DB 4; Length 992;
Best Local Similarity 48.4%; Pred. No. 1.6e-59;
Matches 119; Conservative 50; Mismatches 70; Indels 7; Gaps 3;
QY 1 IDGLRNWLIKPAKSRGRDIVCMRDRVEILAAADHPLSRDNKWWVQKYIETPLLCD 60
Db 411 LDGYQNMWIKPANKCRGRGIILMDNLKKI--LGVVNLISAKSRVYVQKYIERPLILFQ 468
QY 61 TKFDIRQWFLVTDWNPFTIWFYKESYLRFSTQRFSLDKLDSAIHLGNNAVQKYLKNDVGR 120
Db 469 TKFDIRQWFLIINTQPLVWVWFYRESYLRFSSQSEYLSNHHESVHLTNTYAIQKKTNG-KR 527
QY 121 SPFLPAHNNWTSRFOEYLQROGRGAVGSGVIYPSMKKAIAHAMKVAQDHVEPRKNSPEL 180
Db 528 DKRLPSENWDCYCSFOAYLRQIKYKNMMLERIFPGMRKAIVGCMLASQENMDRRPNTFEL 587
QY 181 YGADFVLGRDPRPFLIEINSSPTMPSTPTVTLCAQVQEDTIKYAVDRSCD---IGNF 236
Db 588 FGADFMCNCFPFLIEINSSPDGATTSVTARMCPQCLEDVVKVVIDRRTDPKAE LGNF 647
QY 237 ELLWRQ 242
Db 648 ELAYRQ 653

Search completed: April 4, 2006, 12:51:06
Job time : 78.1431 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:59:01 ; Search time 6.53532 Seconds
(without alignments)
1364.812 Million cell updates/sec

Title: US-10-635-977-2_COPY_73_365

Perfect score: 1560

Sequence: 1 EDIDTSADAVEDLFEAWED.....LCAQVEDTIKVAVDRSCDI 293

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 180808 seqs, 30441898 residues

Total number of hits satisfying chosen parameters: 180808

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA New:*
- 1: /SDSS/ptodata/2/pubppa/US08_NEW_PUB.pdb:*
 - 2: /SDSS/ptodata/2/pubppa/US06_NEW_PUB.pdb:*
 - 3: /SDSS/ptodata/2/pubppa/US07_NEW_PUB.pdb:*
 - 4: /SDSS/ptodata/2/pubppa/PCT_NEW_PUB.pdb:*
 - 5: /SDSS/ptodata/2/pubppa/US05_NEW_PUB.pdb:*
 - 6: /SDSS/ptodata/2/pubppa/US10_NEW_PUB.pdb:*
 - 7: /SDSS/ptodata/2/pubppa/US11_NEW_PUB.pdb:*
 - 8: /SDSS/ptodata/2/pubppa/US60_NEW_PUB.pdb:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	271.5	17.4	592	7	US-11-072-512-3371
2	256.5	16.4	439	7	US-11-072-512-3658
3	180	11.5	404	7	US-11-072-512-3621
4	93	6.0	550	7	US-11-096-568A-12196
5	93	6.0	579	7	US-11-096-568A-12195
6	93	6.0	596	7	US-11-096-568A-12194
7	82	5.3	1137	7	US-11-087-099-5799
8	80	5.1	729	6	US-10-995-561-878
9	80	5.1	744	6	US-10-995-561-876
10	80	5.1	752	6	US-10-995-561-879
11	80	5.1	753	6	US-10-995-561-877
12	79.5	5.1	302	7	US-11-098-686-10927
13	79.5	5.1	433	7	US-11-167-856-26
14	79	5.1	462	7	US-11-096-568A-25994
15	79	5.1	470	7	US-11-096-568A-25993
16	79	5.1	517	7	US-11-096-568A-25992
17	79	5.1	549	7	US-11-085-185-2
18	79	5.1	932	7	US-11-017-550-65
19	79	5.1	932	7	US-11-002-387-65
20	79	5.1	1131	7	US-11-087-099-2209
21	79	5.1	1131	7	US-11-087-099-9522
22	78	5.0	497	6	US-10-793-626-2812
23	77	4.9	993	7	US-11-132-764-7
24	77	4.9	1010	6	US-10-959-611-2
25	76.5	4.9	427	7	US-11-087-099-2254

ALIGNMENTS

RESULT 1

US-11-072-512-3371	427	7	US-11-087-099-12304	Sequence 12304, A
Sequence 3371, Application US/11072512	427	7	US-11-087-099-12358	Sequence 12358, A
Publication No. US20060029945A1	713	6	US-10-995-561-881	Sequence 881, App
GENERAL INFORMATION:	737	6	US-10-995-561-880	Sequence 880, App
APPLICANT: ISOGAI, TAKAO	992	7	US-11-072-512-2473	Sequence 2473, App
APPLICANT: SUGIYAMA, TOMOYASU	992	7	US-11-072-512-2473	Sequence 2473, App
APPLICANT: OTSUKI, TETSUJI	249	6	US-10-926-406A-7	Sequence 7, Appli
APPLICANT: WAKAMATSU, AI	249	6	US-10-926-406A-7	Sequence 7, Appli
APPLICANT: SATO, HIROYUKI	362	7	US-11-072-512-2408	Sequence 2408, App
APPLICANT: ISHII, SHIZUKO	362	7	US-11-072-512-2408	Sequence 2408, App
APPLICANT: YAMAMOTO, JUN-ICHI	1047	7	US-11-167-831-14	Sequence 14, Appli
APPLICANT: ISONO, YUUKO	256	7	US-11-167-831-14	Sequence 14, Appli
APPLICANT: HIO, YURI	179	7	US-11-072-512-2871	Sequence 2871, App
APPLICANT: OTSUKA, KAORU	256	5	US-09-810-501-67	Sequence 67, Appl
APPLICANT: NAGAI, KEIICHI	256	7	US-11-167-831-8	Sequence 8, Appli
APPLICANT: IRIE, RYOTARO	256	7	US-11-167-831-11	Sequence 11, Appl
APPLICANT: TAMECHIKA, ICHIRO	760	6	US-10-858-730-76	Sequence 76, Appl
APPLICANT: SEKI, NAOHICO	1458	7	US-11-096-274-2	Sequence 2, Appli
APPLICANT: YOSHIKAWA, TSUTOMU	2472	7	US-11-203-806A-2	Sequence 2, Appli
APPLICANT: OTSUKA, MOTOYUKI	415	7	US-11-110-851-64	Sequence 64, Appl
APPLICANT: NAGAHARI, KENJI	667	6	US-10-793-626-198	Sequence 198, App
APPLICANT: MASUHO, YASUHIKO	256	7	US-11-167-831-12	Sequence 12, Appl
TITLE OF INVENTION: Novel full length cDNA				
FILE REFERENCE: 084335-0191				
CURRENT APPLICATION NUMBER: US/11/072,512				
CURRENT FILING DATE: 2005-03-07				
PRIOR APPLICATION NUMBER: US 60/350,978				
PRIOR FILING DATE: 2002-01-25				
PRIOR APPLICATION NUMBER: JP 2001-379298				
PRIOR FILING DATE: 2001-11-05				
NUMBER OF SEQ ID NOS: 4096				
SOFTWARE: PatentIn Ver. 2.1				
SEQ ID NO 3371				
LENGTH: 592				
TYPE: PRT				
ORGANISM: Homo sapiens				
US-11-072-512-3371				

Query Match	17.4%	Score 271.5	DB 7	Length 592
Best Local Similarity	29.8%	Pred. No. 1.1e-18		
Matches	72	Conservative 43	Mismatches 80	Indels 47
Gaps	8			
QY	41	YFSOQALLNRTTSVNPQTDIDLRNIWIIPAKSRGRDVCMDRVSEILELAADHPL	100	
DB	195	YFQERQMLGTK-----HSYWCIPAEUSRGHILFSDFKDFI-----	232	
QY	101	SRDNKVVQVYETPLLICDTKFDIQRWFLVTDWNPITWIFYKESYLRFSTQRFSDKLI-	159	

Db 233 -FDMIVQKYNPLIGRYKCDLRIYVCTGFKPLTIIVYQGLVRATEKFDLSNLQ 291
QY 160 DSAIHLNNAVOK-----YLKNDVGRSPILLPAHNMWTSRFOEYLQROGRGAW--GS 210
Db 292 NYAHLTNSINKSGASYEIKIVEIGHG-----CKWTLRSRFPFYLRS-----WDVDDL 339
QY 211 NYPSMKATAHAMKVAQDHVEPRKNSFELYGADFVLGRDPRFWLIEINSSP--TWHPS 268
Db 340 LLWKKIHRMVLITLAIAPSPVFAANCFLFGFDLILDDNKLKPLLEVNYSPLTLD 399
QY 269 PV 270
Db 400 DV 401

RESULT 2

US-11-072-512-3658
; Sequence 3658, Application US/11072512
; Publication No. US20060029945A1

GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOUYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cdna
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3658
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Homo sapiens

US-11-072-512-3658

Query Match 16.4%; Score 256.5; DB 7; Length 439;
Best Local Similarity 25.5%; Pred. No. 2.1e-17;
Matches 84; Conservative 56; Mismatches 109; Indels 81; Gaps 13;

QY 10 VEDLTEAEWE-----DLT-----QOYSLVHGDAFISNRYFSQC 45
Db 65 VEVKDEGEWDFYWCDSWLNRENFDHTYNDHVRISHFNHYELTRKNYMNLRKFRKL 124
QY 46 QALINRITSVN---POT-----DIDGLRN-----IWIIPAAKSRGRDIVCM 88
Db 125 EREAGKLEAAKCDFFPKTFEMPCRYHLFVEFRKNPGITWIMKPVARSQKGIPLRRLK 184
QY 89 EILEAADAHLPSRDNK-----WVQKYIETPLLCDTKFDIRQWFLVDWNP 141
Db 185 DIVDW-RKDRSSDDQKDDIPVENYVAQRYTENFYLIIGGRKFDLRVYVLMS-----V 236
QY 142 YKESYLFSTORFSLDKLSDSAIHLNNAVOKYLNKNDVGRSPLLPAHN-----MWTSTRFQ 196

Db 237 FABCLLWSGHR-----QDVHLTNVAVOK-----TSPDYHPKKGCKWTQOR 279
QY 197 EYL-QROGRGAVGWSVVIYPSMKKAIAHAMKVAQDHVEPRKNSFELYCADFVLGRDPRFWL 255
Db 280 QYLASHKHGPEAV--ETLFRDIDNIFVKSLSQSVQKVIISDKHCFELYGYDILIDODLKPWL 337
QY 256 IENSSPPTWHPSPPTVTAQLCAQVOEDTIKV 285
Db 338 LEVNASPSLTASSQEDYELKTCLEDTLHV 367

RESULT 3

US-11-072-512-3621

; Sequence 3621, Application US/11072512
; Publication No. US20060029945A1

GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOUYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cdna
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3621
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Homo sapiens

US-11-072-512-3621

Query Match 11.5%; Score 180; DB 7; Length 404;
Best Local Similarity 25.6%; Pred. No. 6.8e-10;
Matches 63; Conservative 33; Mismatches 84; Indels 66; Gaps 12;

QY 4 DTSADAVEDLTEAEWEDLTQOYSLVHGDAFISNRYFSQCQALLNRITSVNPQT---D 60
Db 183 DAAAPAEDEL---PW-----TSPGYLRPQVRL--RMEFFETETRLD 219
QY 61 IDGLR-----NIWIIKPAKSRGRDIVCMRVEEILELAA-----ADHPLSRDNK 105
Db 220 LKHEREAFFTLFDETQIWKPTASNOQKGIFLLRNQEVAALQAKTRSMEDDPHHTKP 279
QY 106 W-----WVQKYIETPLLCDTKFDIRQWFLVDWNP 159
Db 280 FRGQARVQORYIQNPLLVGRKFDVRSYLLIACCTPYMI--FFGHGYARLTSLYDPHSS 338
QY 160 DSAIHLNNAVOKYLNKNDVGRSP---LLPAHNMWTSRFOEYL-----QROGRGAVGWSV 211
Db 339 DLGHLTNQFMOK-----KSPLYMLLKHTVWSMEHLNRYISDTFWKARGLAKDW--- 388
QY 212 IYPSMK 217
Db 389 VFTTLK 394

RESULT 4
US-11-096-568A-12196
; Sequence 12196, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 12196
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(550)
; OTHER INFORMATION: Ceres Seq. ID no. 15220180
US-11-096-568A-12196

Query Match 6.0%; Score 93; DB 7; Length 550;
Best Local Similarity 20.9%; Pred. No. 0.4;
Matches 55; Conservative 46; Mismatches 98; Indels 64; Gaps 11;
QY 39 RNYFSQCQALLNRI-----TSVNPQTDIDGLRNIIWIKPAA---KSRGRDIV---CMDRV 87
DB 314 RNLFRKCANLIERLNSSPDSVNPRLMHALRELEMDSPKSSSENEESGRLITPQACNRI 373
QY 88 BEILELAAADHPLSRDNKVVQKYIETPLLICDTKPD--IRQWFLVTDWNPPLTIWPKES 145
DB 374 AEVQEAVRKVELVAEERKMGLYKARTAVEACDRELDEKARQ-----VQEFKAE 422
QY 146 YLRFSTQRFSLDKLDSAIHLCCNNAVQKY-LKNDVGR--SPLLPAHNMTSTRFQ----- 197
DB 423 RLRRKQKQ---VEELESIVRLKQAEAEFOLKASEARQAEARLQSIATLAKSERAEOQYASL 479
QY 198 YLQROGRGAVMGSVIYPSMKKAIHAAMKVAQDHVEPRKNSFELYGADFVLGRDRFPWLIE 257
DB 480 YLKRRLKEA-----EAEKQFLFEKIKLQDGHRRPQASS----- 512
QY 258 INSSPTMHPSTPTVTAQLCAQVOE 280
DB 513 --SVPADSSQAPSQALMLSKIQD 533

RESULT 5
US-11-096-568A-12195
; Sequence 12195, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 12195
; LENGTH: 579
; TYPE: PRT
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(579)
; OTHER INFORMATION: Ceres Seq. ID no. 15220179
US-11-096-568A-12195

Query Match 6.0%; Score 93; DB 7; Length 579;
Best Local Similarity 20.9%; Pred. No. 0.43;

Matches 55; Conservative 46; Mismatches 98; Indels 64; Gaps 11;
QY 39 RNYFSQCQALLNRI-----TSVNPQTDIDGLRNIIWIKPAA---KSRGRDIV---CMDRV 87
DB 343 RNLFRKCANLIERLNSSPDSVNPRLMHALRELEMDSPKSSSENEESGRLITPQACNRI 402
QY 88 BEILELAAADHPLSRDNKVVQKYIETPLLICDTKPD--IRQWFLVTDWNPPLTIWPKES 145
DB 403 AEVQEAVRKVELVAEERKMGLYKARTAVEACDRELDEKARQ-----VQEFKAE 451
QY 146 YLRFSTQRFSLDKLDSAIHLCCNNAVQKY-LKNDVGR--SPLLPAHNMTSTRFQ----- 197
DB 452 RLRRKQKQ---VEELESIVRLKQAEAEFOLKASEARQAEARLQSIATLAKSERAEOQYASL 508
QY 198 YLQROGRGAVMGSVIYPSMKKAIHAAMKVAQDHVEPRKNSFELYGADFVLGRDRFPWLIE 257
DB 509 YLKRRLKEA-----EAEKQFLFEKIKLQDGHRRPQASS----- 541
QY 258 INSSPTMHPSTPTVTAQLCAQVOE 280
DB 542 --SVPADSSQAPSQALMLSKIQD 562

RESULT 6
US-11-096-568A-12194
; Sequence 12194, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptide
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 12194
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(596)
; OTHER INFORMATION: Ceres Seq. ID no. 15220178
US-11-096-568A-12194

Query Match 6.0%; Score 93; DB 7; Length 596;
Best Local Similarity 20.9%; Pred. No. 0.44;
Matches 55; Conservative 46; Mismatches 98; Indels 64; Gaps 11;
QY 39 RNYFSQCQALLNRI-----TSVNPQTDIDGLRNIIWIKPAA---KSRGRDIV---CMDRV 87
DB 360 RNLFRKCANLIERLNSSPDSVNPRLMHALRELEMDSPKSSSENEESGRLITPQACNRI 419
QY 88 BEILELAAADHPLSRDNKVVQKYIETPLLICDTKPD--IRQWFLVTDWNPPLTIWPKES 145
DB 420 AEVQEAVRKVELVAEERKMGLYKARTAVEACDRELDEKARQ-----VQEFKAE 468
QY 146 YLRFSTQRFSLDKLDSAIHLCCNNAVQKY-LKNDVGR--SPLLPAHNMTSTRFQ----- 197
DB 469 RLRRKQKQ---VEELESIVRLKQAEAEFOLKASEARQAEARLQSIATLAKSERAEOQYASL 525
QY 198 YLQROGRGAVMGSVIYPSMKKAIHAAMKVAQDHVEPRKNSFELYGADFVLGRDRFPWLIE 257
DB 526 YLKRRLKEA-----EAEKQFLFEKIKLQDGHRRPQASS----- 558
QY 258 INSSPTMHPSTPTVTAQLCAQVOE 280
DB 559 --SVPADSSQAPSQALMLSKIQD 579

RESULT 7
US-11-087-099-5799
; Sequence 5799, Application US/11087099

```
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 5799
; LENGTH: 1137
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1137)
; OTHER INFORMATION: unsure at all Xaa locations
US-11-087-099-5799

Query Match 5.1%; Score 82; DB 7; Length 1137;
Best Local Similarity 18.9%; Pred. No. 13;
Matches 41; Conservative 39; Mismatches 65; Indels 72; Gaps 8;

QY 43 SOQALLNRITSVNPQTDIDGLRNIIK-----PAAKSRGRDIVCM-----DRVVEE 89
Db 631 SEIVRLIDTATVILAVDVGGLVGNWIKIAELTGLPIGEATGKHLTLVEDSSIDRVKK 690
QY 90 ILELAADHPLSRDNKVVQKIETPLLCIDTKFDIRQWFLVTDWNPPLTIWFKESYLRF 149
Db 691 MLNLAL-----LGBEEK-----NVQFEIKTHGSKMDSGPISL-----V 723
QY 150 STQSFSLDKLDSAIHLC-----NNAVQYKLNK-----VGRSPLL----- 184
Db 724 VNACASGRDRDNVVGCVFVAHDITAKNVMDKFIIEGDKYKAVQNRNPLPIPIFGTDEF 783
QY 185 -----PAHNMWTSRFOEYLQROGRGAVGWGSVIYP 214
Db 784 GWCEWNPAMKLTGWKKEEVDKMLLGEIFGTQMAP 820

RESULT 8
US-10-995-561-878
; Sequence 878, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 878
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-878

Query Match 5.1%; Score 80; DB 6; Length 729;
Best Local Similarity 19.2%; Pred. No. 11;
Matches 51; Conservative 42; Mismatches 97; Indels 76; Gaps 11;

QY 43 SOQALLNRITSVNPQTDIDGLRNIIKPAAKSRGRDIVCMRVEIFELELAADHPLSR 102
Db 278 TDCENLLKRFVLNP-----IK-----RG-----TLEQIMK----- 303
QY 103 DNKWV-----VQKIETPLLCIDTK-----FDIRQWFLVTDWNPPLTIWFY 142
Db 304 -DRWINAGHEDELKPFVEPELDSIQKRIDIMVGMGYSQEEISLKMKYDEIT----- 358
QY 143 KESYLRFSTQRFSLDKLDSA-----IHLCNNAVQYKLNKDVGRSPLLPAHNMWTSRFOEY 198
Db 359 -ATYLLGRKSSSELDASDSSSSNLSLAKVPSDDLNNSTGQSPHHKVQSVSSSQKRR 417
QY 199 LQROGRGAVGWGSVIYPSMKKAIHAHMKVAQDHVEPRKNSFELYGADFVLGRDFRPMLEI 258
Db 418 YSDHAGPAIPSVVAYPKRSQTSADSLKEDGISRRKSSGSVAVG-----GKGIAP----- 467
QY 259 NSPPTM-HPSTPTVTAOLCAQVQEDTI 283
Db 468 -ASPMGLGNASNPKNKADIPERKKSSTV 492

RESULT 9
US-10-995-561-876
; Sequence 876, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 876
; LENGTH: 744
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-876

Query Match 5.1%; Score 80; DB 6; Length 744;
Best Local Similarity 19.2%; Pred. No. 11;
Matches 51; Conservative 42; Mismatches 97; Indels 76; Gaps 11;

QY 43 SOQALLNRITSVNPQTDIDGLRNIIKPAAKSRGRDIVCMRVEIFELELAADHPLSR 102
Db 278 TDCENLLKRFVLNP-----IK-----RG-----TLEQIMK----- 303
QY 103 DNKWV-----VQKIETPLLCIDTK-----FDIRQWFLVTDWNPPLTIWFY 142
Db 304 -DRWINAGHEDELKPFVEPELDSIQKRIDIMVGMGYSQEEISLKMKYDEIT----- 358
QY 143 KESYLRFSTQRFSLDKLDSA-----IHLCNNAVQYKLNKDVGRSPLLPAHNMWTSRFOEY 198
Db 359 -ATYLLGRKSSSELDASDSSSSNLSLAKVPSDDLNNSTGQSPHHKVQSVSSSQKRR 417
QY 199 LQROGRGAVGWGSVIYPSMKKAIHAHMKVAQDHVEPRKNSFELYGADFVLGRDFRPMLEI 258
Db 418 YSDHAGPAIPSVVAYPKRSQTSADSLKEDGISRRKSSGSVAVG-----GKGIAP----- 467
QY 259 NSPPTM-HPSTPTVTAOLCAQVQEDTI 283
Db 468 -ASPMGLGNASNPKNKADIPERKKSSTV 492

RESULT 10
US-10-995-561-879
; Sequence 879, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 879
; LENGTH: 752
; TYPE: PRT
US-10-995-561-879

Query Match 5.1%; Score 80; DB 6; Length 729;
Best Local Similarity 19.2%; Pred. No. 11;
Matches 51; Conservative 42; Mismatches 97; Indels 76; Gaps 11;

QY 43 SOQALLNRITSVNPQTDIDGLRNIIKPAAKSRGRDIVCMRVEIFELELAADHPLSR 102
Db 278 TDCENLLKRFVLNP-----IK-----RG-----TLEQIMK----- 303
QY 103 DNKWV-----VQKIETPLLCIDTK-----FDIRQWFLVTDWNPPLTIWFY 142
Db 304 -DRWINAGHEDELKPFVEPELDSIQKRIDIMVGMGYSQEEISLKMKYDEIT----- 358
QY 143 KESYLRFSTQRFSLDKLDSA-----IHLCNNAVQYKLNKDVGRSPLLPAHNMWTSRFOEY 198
```


GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: April 4, 2006, 12:55:16 ; Search time 42.2072 Seconds
(without alignments)
2900.544 Million cell updates/sec
Title: US-10-635-977-2_COPY_73_365
Perfect score: 1560
Sequence: 1 EDIDTSADAVDELTEAEWED.....LCAQVQEDTIKVAVDRSCDI 293
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA Main:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1560	100.0	293	4	US-10-615-659-14
2	1560	100.0	293	4	US-10-635-977-14
3	1560	100.0	541	4	US-10-615-659-2
4	1560	100.0	541	4	US-10-615-659-13
5	1560	100.0	541	4	US-10-635-977-2
6	1560	100.0	541	4	US-10-635-977-13
7	1247	79.9	242	4	US-10-615-659-4
8	1247	79.9	242	4	US-10-635-977-4
9	908	58.2	352	4	US-10-615-659-7
10	908	58.2	352	4	US-10-635-977-7
11	908	58.2	352	5	US-10-756-149-5267
12	908	58.2	488	4	US-10-275-595A-5
13	905	58.0	326	4	US-10-108-260A-4209
14	897	57.5	362	3	US-09-928-298-617
15	897	57.5	362	4	US-10-102-806-617
16	817	52.4	292	4	US-10-615-659-5
17	817	52.4	292	4	US-10-635-977-5
18	636	40.8	992	6	US-11-097-143-23727
19	515.5	33.0	719	6	US-11-097-143-23415
20	374	24.0	566	5	US-10-450-763-36330
21	324	20.8	496	6	US-11-097-143-19014
22	321.5	20.6	330	4	US-10-615-659-26
23	321.5	20.6	330	4	US-10-635-977-26
24	297.5	19.1	827	6	US-11-097-143-9312
25	296	19.0	1226	5	US-10-756-149-5184
26	290.5	18.6	989	6	US-11-097-143-7680
27	287	18.4	423	4	US-10-615-659-8

28	287	18.4	423	4	US-10-635-977-8	Sequence 8, Appli
29	284	18.2	268	4	US-10-424-599-205823	Sequence 205823,
30	274	17.6	49	4	US-10-615-659-21	Sequence 21, Appl
31	274	17.6	49	4	US-10-615-659-22	Sequence 22, Appl
32	274	17.6	49	4	US-10-635-977-21	Sequence 21, Appl
33	274	17.6	49	4	US-10-635-977-22	Sequence 22, Appl
34	272	17.4	487	6	US-11-097-143-4080	Sequence 4080, Ap
35	271.5	17.4	524	3	US-09-864-761-38213	Sequence 38213, A
36	271.5	17.4	592	4	US-10-104-047-3371	Sequence 3371, A
37	270	17.3	917	6	US-11-097-143-21303	Sequence 21303, A
38	258	16.4	92	4	US-10-424-599-262294	Sequence 262294,
39	256.5	16.4	439	4	US-10-104-047-3658	Sequence 3658, Ap
40	255.5	16.4	379	4	US-10-615-659-6	Sequence 6, Appli
41	255.5	16.4	379	4	US-10-635-977-6	Sequence 6, Appli
42	250.5	16.1	377	4	US-10-210-130-118	Sequence 118, App
43	250.5	16.1	377	4	US-10-250-613-1	Sequence 1, Appli
44	208	13.3	720	6	US-11-097-143-12744	Sequence 12744, A
45	205.5	13.2	553	4	US-10-108-260A-2495	Sequence 2495, Ap

ALIGNMENTS

RESULT 1

US-10-615-659-14
; Sequence 14, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-14

Query Match	100.0%	Score 1560;	DB 4;	Length 293;
Best Local Similarity	100.0%;	Pred. No. 2.7e-151;		
Matches 293;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	EDIDTSADAVDELTEAEWEDLTQQYISLVHGDAFISNSRNYFSQCOALLNRITSVNPQTD	60	
Db	1	EDIDTSADAVDELTEAEWEDLTQQYISLVHGDAFISNSRNYFSQCOALLNRITSVNPQTD	60	
Qy	61	IDGLRNIIIKPAKSRGRDIVCMRVVEIELEAAADHPLSRDNKWWVQKYTELLICD	120	
Db	61	IDGLRNIIIKPAKSRGRDIVCMRVVEIELEAAADHPLSRDNKWWVQKYTELLICD	120	
Qy	121	TKFDIRQWFLVTDNPLTIWPKYSYLRFTQSRSLDKLSAHLCCNNAVQKYLKNDVGR	180	
Db	121	TKFDIRQWFLVTDNPLTIWPKYSYLRFTQSRSLDKLSAHLCCNNAVQKYLKNDVGR	180	
Qy	181	SPLLPAHNMWTSRFOEYLRQGRGAVMGSVIYFSMKKAIHAHMKVAQDHVEPRKNSFEL	240	
Db	181	SPLLPAHNMWTSRFOEYLRQGRGAVMGSVIYFSMKKAIHAHMKVAQDHVEPRKNSFEL	240	
Qy	241	YGADFVLGRDRPRPLIEINSSPTMHPSTPTVAQLCAQVQEDTIKVAVDRSCDI	293	
Db	241	YGADFVLGRDRPRPLIEINSSPTMHPSTPTVAQLCAQVQEDTIKVAVDRSCDI	293	

RESULT 2

US-10-635-977-14
; Sequence 14, Application US/10635977
; Publication No. US20040171131A1

; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-14

Query Match 100.0%; Score 1560; DB 4; Length 293;
Best Local Similarity 100.0%; Pred. No. 2.7e-151;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 EDITSDADAVEDLTEAEWEDLTQYYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 60
Db 1 EDITSDADAVEDLTEAEWEDLTQYYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 60
Qy 61 IDGLRNIWIIKPAAKSRGRDIVCMDRVVEILEAAADHPLSRDNKWKVQKIETPLLICD 120
Db 61 IDGLRNIWIIKPAAKSRGRDIVCMDRVVEILEAAADHPLSRDNKWKVQKIETPLLICD 120
Qy 121 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 180
Db 121 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 180
Qy 181 SP LLPAHNNMTSTRFOEYLQROGRGAVGWSVIYPSMKKATAHAMKVAQDHVEPRKNSFEL 240
Db 181 SP LLPAHNNMTSTRFOEYLQROGRGAVGWSVIYPSMKKATAHAMKVAQDHVEPRKNSFEL 240
Qy 241 YGADFVLGRDRFRPWLIEINSSPTMHPSTPVTQAQLCAQVQEDTIKVAVDRSCDI 293
Db 241 YGADFVLGRDRFRPWLIEINSSPTMHPSTPVTQAQLCAQVQEDTIKVAVDRSCDI 293

RESULT 3
US-10-615-659-2
; Sequence 2, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-2

Query Match 100.0%; Score 1560; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 6.3e-151;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 EDITSDADAVEDLTEAEWEDLTQYYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 60
Db 73 EDITSDADAVEDLTEAEWEDLTQYYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 132

Qy 61 IDGLRNIWIIKPAAKSRGRDIVCMDRVVEILEAAADHPLSRDNKWKVQKIETPLLICD 120
Db 133 IDGLRNIWIIKPAAKSRGRDIVCMDRVVEILEAAADHPLSRDNKWKVQKIETPLLICD 192
Qy 121 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 180
Db 193 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 252
Qy 181 SP LLPAHNNMTSTRFOEYLQROGRGAVGWSVIYPSMKKATAHAMKVAQDHVEPRKNSFEL 240
Db 253 SP LLPAHNNMTSTRFOEYLQROGRGAVGWSVIYPSMKKATAHAMKVAQDHVEPRKNSFEL 312
Qy 241 YGADFVLGRDRFRPWLIEINSSPTMHPSTPVTQAQLCAQVQEDTIKVAVDRSCDI 293
Db 313 YGADFVLGRDRFRPWLIEINSSPTMHPSTPVTQAQLCAQVQEDTIKVAVDRSCDI 365

RESULT 4
US-10-615-659-13
; Sequence 13, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-13

Query Match 100.0%; Score 1560; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 6.3e-151;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 EDITSDADAVEDLTEAEWEDLTQYYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 60
Db 73 EDITSDADAVEDLTEAEWEDLTQYYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 132
Qy 61 IDGLRNIWIIKPAAKSRGRDIVCMDRVVEILEAAADHPLSRDNKWKVQKIETPLLICD 120
Db 133 IDGLRNIWIIKPAAKSRGRDIVCMDRVVEILEAAADHPLSRDNKWKVQKIETPLLICD 192
Qy 121 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 180
Db 193 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 252
Qy 181 SP LLPAHNNMTSTRFOEYLQROGRGAVGWSVIYPSMKKATAHAMKVAQDHVEPRKNSFEL 240
Db 253 SP LLPAHNNMTSTRFOEYLQROGRGAVGWSVIYPSMKKATAHAMKVAQDHVEPRKNSFEL 312
Qy 241 YGADFVLGRDRFRPWLIEINSSPTMHPSTPVTQAQLCAQVQEDTIKVAVDRSCDI 293
Db 313 YGADFVLGRDRFRPWLIEINSSPTMHPSTPVTQAQLCAQVQEDTIKVAVDRSCDI 365

RESULT 5
US-10-635-977-2
; Sequence 2, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977

; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-2

Query Match 100.0%; Score 1560; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 6.3e-151;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EDIDTSADAVEDLTEAEWEDLTQQYSLVHGDAFISNRNYSQOALLNRITSVNPQTD 60
DB 73 EDIDTSADAVEDLTEAEWEDLTQQYSLVHGDAFISNRNYSQOALLNRITSVNPQTD 132
QY 61 IDGLRNIIWIKPAAKSRGRDIVCMRDVVEILELAADHPLSRDNKWWVQKVIETPLLICD 120
DB 133 IDGLRNIIWIKPAAKSRGRDIVCMRDVVEILELAADHPLSRDNKWWVQKVIETPLLICD 192
QY 121 TKFDIRQWFLVTDNPLTIWFKESYLRPSTQRFSLDKLDSAIHLCLNNAVQKYLKNDVGR 180
DB 193 TKFDIRQWFLVTDNPLTIWFKESYLRPSTQRFSLDKLDSAIHLCLNNAVQKYLKNDVGR 252
QY 181 SPLLPAHNNMTSTRFOEYLQRGAVGWSVIYPSMKKAIHAAMKVAQDHVEPRKNSPEL 240
DB 253 SPLLPAHNNMTSTRFOEYLQRGAVGWSVIYPSMKKAIHAAMKVAQDHVEPRKNSPEL 312
QY 241 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDKSCDI 293
DB 313 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDKSCDI 365

RESULT 6
US-10-635-977-13
; Sequence 13, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR FILING DATE: 2002-07-09
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-13

Query Match 100.0%; Score 1560; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 6.3e-151;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EDIDTSADAVEDLTEAEWEDLTQQYSLVHGDAFISNRNYSQOALLNRITSVNPQTD 60
DB 73 EDIDTSADAVEDLTEAEWEDLTQQYSLVHGDAFISNRNYSQOALLNRITSVNPQTD 132
QY 61 IDGLRNIIWIKPAAKSRGRDIVCMRDVVEILELAADHPLSRDNKWWVQKVIETPLLICD 120
DB 133 IDGLRNIIWIKPAAKSRGRDIVCMRDVVEILELAADHPLSRDNKWWVQKVIETPLLICD 192

QY 121 TKFDIRQWFLVTDNPLTIWFKESYLRPSTQRFSLDKLDSAIHLCLNNAVQKYLKNDVGR 180
DB 193 TKFDIRQWFLVTDNPLTIWFKESYLRPSTQRFSLDKLDSAIHLCLNNAVQKYLKNDVGR 252
QY 181 SPLLPAHNNMTSTRFOEYLQRGAVGWSVIYPSMKKAIHAAMKVAQDHVEPRKNSPEL 240
DB 253 SPLLPAHNNMTSTRFOEYLQRGAVGWSVIYPSMKKAIHAAMKVAQDHVEPRKNSPEL 312
QY 241 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDKSCDI 293
DB 313 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDKSCDI 365

RESULT 7
US-10-615-659-4
; Sequence 4, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-4

Query Match 79.9%; Score 1247; DB 4; Length 242;
Best Local Similarity 100.0%; Pred. No. 2.8e-119;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 IDGLRNIIWIKPAAKSRGRDIVCMRDVVEILELAADHPLSRDNKWWVQKVIETPLLICD 120
DB 1 IDGLRNIIWIKPAAKSRGRDIVCMRDVVEILELAADHPLSRDNKWWVQKVIETPLLICD 60
QY 121 TKFDIRQWFLVTDNPLTIWFKESYLRPSTQRFSLDKLDSAIHLCLNNAVQKYLKNDVGR 180
DB 61 TKFDIRQWFLVTDNPLTIWFKESYLRPSTQRFSLDKLDSAIHLCLNNAVQKYLKNDVGR 120
QY 181 SPLLPAHNNMTSTRFOEYLQRGAVGWSVIYPSMKKAIHAAMKVAQDHVEPRKNSPEL 240
DB 121 SPLLPAHNNMTSTRFOEYLQRGAVGWSVIYPSMKKAIHAAMKVAQDHVEPRKNSPEL 180
QY 241 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDKSCDI 293
DB 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDKSCDI 233

RESULT 8
US-10-635-977-4
; Sequence 4, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR FILING DATE: 2002-07-09
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4

Db 2 DIDKDLAPLYLTPEGWSLFLQRYQVYVHGAELRHLDLTQVQRCEDILOQLQAVVPQIDM 61
QY 62 DGLRNIIWIKPAKSGRGRDIVCMRVBEILEAAADHPLSRDNKWWVQKYIETPLLCDT 121
Db 62 EGDRIWIVKPGAKSGRGIMCWDHLEMLKLVNGNPPVMDGKVVQKYIERPLLIFFT 121
QY 122 KFDIRQWFLVTDWNPITWIFYKESYLRFSRSLDKLDSAIHLGNNAVQKYLKNDVGRS 181
Db 122 KFDLRQWFLVTDWNPITWIFYKESYLRFSRSLDKLDSAIHLGNNAVQKYLKNDVGRS 181
QY 182 PLLPAHNNMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVPRKNSFELY 241
Db 182 PLLPPDNWSSQRFQAHLEQMGAPNAWSTIIVPGMKDAVIHALQTSQDVTQCRKASFELY 241
QY 242 GADFVLGRDRPWLIEINSSPTMHPSTPVTQAQLCAQVQEDTIKVAVDKSD 292
Db 242 GADFVFGEDFQWPLIEINASPTMAPSTAVTARLCAGVQADTLRVVIDRMLD 292

RESULT 12

US-10-275-595A-5

; Sequence 5, Application US/10275595A

; Publication No. US20040078804A1

; GENERAL INFORMATION:

; APPLICANT: YUE, Henry

; APPLICANT: TANG, Y. Tom

; APPLICANT: AU-YOUNG, Janice

; APPLICANT: LU, Dying Aina M.

; APPLICANT: BAUGHN, Mariah R.

; APPLICANT: HILLMAN, Jennifer L.

; APPLICANT: AZIMZAI, Yalda

; APPLICANT: LAL, Preeti

; APPLICANT: YAO, Monique G.

; APPLICANT: BANDMAN, Olga

; APPLICANT: BURFORD, Neil

; APPLICANT: BATEA, Sajeev

; APPLICANT: KEARNEY, Liam

; APPLICANT: POLICKY, Jennifer L.

; TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS

; FILE REFERENCE: PF-0772 USN

; CURRENT APPLICATION NUMBER: US/10/275,595A

; CURRENT FILING DATE: 2003-06-13

; PRIOR APPLICATION NUMBER: US 60/201,960

; PRIOR FILING DATE: 2000-05-05

; PRIOR APPLICATION NUMBER: US 60/202,729

; PRIOR FILING DATE: 2000-05-08

; PRIOR APPLICATION NUMBER: US 60/209,705

; PRIOR FILING DATE: 2000-06-05

; PRIOR APPLICATION NUMBER: US 60/210,149

; PRIOR FILING DATE: 2000-06-07

; PRIOR APPLICATION NUMBER: US 60/213,215

; PRIOR FILING DATE: 2000-06-21

; NUMBER OF SEQ ID NOS: 68

; SOFTWARE: PERL Program

; SEQ ID NO 5

; LENGTH: 488

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; OTHER INFORMATION: Incyte ID No: 2156553CD1

US-10-275-595A-5

Query Match 58.2%; Score 908; DB 4; Length 488;

Best Local Similarity 56.7%; Pred. No. 4.9e-84;

Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;

QY 2 DIDTSADAVEDLTEAEWEDLTQQYYSLVHGDAFISNRNYSQCALNRITSVNPQTDI 61
Db 2 DIDKDLAPLYLTPEGWSLFLQRYQVYVHGAELRHLDLTQVQRCEDILOQLQAVVPQIDM 61
QY 62 DGLRNIIWIKPAKSGRGRDIVCMRVBEILEAAADHPLSRDNKWWVQKYIETPLLCDT 121

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 617
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (307)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-298-617

Query Match 57.5%; Score 897; DB 3; Length 362;
Best Local Similarity 57.5%; Pred. No. 4.4e-83;
Matches 161; Conservative 47; Mismatches 72; Indels 0; Gaps 0;

QY	13	LTEAEWEDLTQQYYSLVHGDAFISNSRNYFSQCALNLRITSVNPQTDIDGLRNIIKIP	72
Db	30	LTPEGWSLFLQRYQVVHGAELRHLDTVQRCEDILQQQAVVPQIDMEGDRNIWIKP	89
QY	73	AAKSRGRDIVCMRDRVEILELAADHPLSRDNKVVQKYYIETPLLICDTKFDIRQWFLVT	132
Db	90	GAKSRGRGIMCNDHLEEMKLKLVNGNPVVMKDGKVVQKYIERPLLIIFGTFKFDLRQWFLVT	149
QY	133	DWNPLTIWFYKESYLRFSQSLDKLSAIHLKNNVAVQKYLKNDVGRSPLLPAHNMWTS	192
Db	150	DWNPLTVWFYRDSYIRFSTQPFSLKNLDSNVHLKNNISIQKHLENSCHRHPLLPDPNNWSS	209
QY	193	TRFQEVLRQGRGAVNGSVIYPSMKKAIHAHAKVAQDHYEPRKNSFELYGADFVLGRDPR	252
Db	210	ORFQAHLOENGAPNASTIIVPGMKDAVIHALQTSQDTVQCRKASFELYGADVFVGEDFQ	269
QY	253	PWLIENSPTMHPSTPVTQAQLCAQVOEDTIKVAVDNRSCD	292
Db	270	PWLIENASPTWAPSTAVTARLCAGVQADTLRVVIDRXLD	309

RESULT 15

US-10-102-806-617
; Sequence 617, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103P1C1
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 617
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (307)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-102-806-617

Query Match 57.5%; Score 897; DB 4; Length 362;
Best Local Similarity 57.5%; Pred. No. 4.4e-83;
Matches 161; Conservative 47; Mismatches 72; Indels 0; Gaps 0;

QY	13	LTEAEWEDLTQQYYSLVHGDAFISNSRNYFSQCALNLRITSVNPQTDIDGLRNIIKIP	72
Db	30	LTPEGWSLFLQRYQVVHGAELRHLDTVQRCEDILQQQAVVPQIDMEGDRNIWIKP	89
QY	73	AAKSRGRDIVCMRDRVEILELAADHPLSRDNKVVQKYYIETPLLICDTKFDIRQWFLVT	132

Db	90	GAKSRGRGIMCNDHLEEMKLKLVNGNPVVMKDGKVVQKYIERPLLIIFGTFKFDLRQWFLVT	149
QY	133	DWNPLTIWFYKESYLRFSQSLDKLSAIHLKNNVAVQKYLKNDVGRSPLLPAHNMWTS	192
Db	150	DWNPLTVWFYRDSYIRFSTQPFSLKNLDSNVHLKNNISIQKHLENSCHRHPLLPDPNNWSS	209
QY	193	TRFQEVLRQGRGAVNGSVIYPSMKKAIHAHAKVAQDHYEPRKNSFELYGADFVLGRDPR	252
Db	210	ORFQAHLOENGAPNASTIIVPGMKDAVIHALQTSQDTVQCRKASFELYGADVFVGEDFQ	269
QY	253	PWLIENSPTMHPSTPVTQAQLCAQVOEDTIKVAVDNRSCD	292
Db	270	PWLIENASPTWAPSTAVTARLCAGVQADTLRVVIDRXLD	309

Search completed: April 4, 2006, 12:58:00
Job time : 43.2072 secs


```
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3658
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3658

Query Match 16.4%; Score 256.5; DB 2; Length 439;
Best Local Similarity 25.5%; Pred. No. 8.9e-20;
Matches 84; Conservative 56; Mismatches 109; Indels 81; Gaps 13;

QY 10 VEDLTAEWB-----DLT-----QYYSLVHGDAFISNSRYFSQC 45
Db 65 VEVKDEGEWFYWCDSWLENFDHTYMDHEVRIASHFRNHVELTRKNYMWKNLKRKQL 124
QY 46 QALLNRITSVN-----POT-----DIIDLGN-----IWIIPAAKSRGRDIVCMRDVE 88
Db 125 ERAGKLEAAKCDFFPKTFEMPCEYHLFVEEPRKNPGITWIMKPVARSQKGIFLFRLLX 184
QY 89 EILELAADHPLSRDNK-----WVQYIETPLLICDTKFDIROWFLVTDWNPITWIF 141
Db 185 DIVDM-RKOTRSDQDKDDIPVENYAQRYIENPYLIGGRKFDLRVVVLVMS-----V 236
QY 142 YKESYLRPSTQFSLDKLSAIHLCNNAVQKYLKNDVGRSPLLPAHN-----MWTSTRFQ 196
Db 237 FAECLLWSGHR-----QDVHLTNVAVQK-----TSPDHPKKGCKWTQORFR 279
QY 197 EYL-ORQGRGAVGMSVTPYMKKAIAHAMKVAQDHPVEPRKNSPELYGADFVLGRDFPWL 255
Db 280 QYLASKHGPAV--ETLFRIDINIFVKSLSQSVQKVIISDKHCFELYGVDILIDQDLKPL 337
QY 256 IEINSSPTMHPSTPVTAQLCAQVQEDTIKV 285
Db 338 LEVNASPLTASSQEDYELKTCLEDTLHV 367

RESULT 3
US-09-270-767-46703
; Sequence 46703, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46703
; LENGTH: 561
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-46703

Query Match 13.6%; Score 212; DB 2; Length 561;
Best Local Similarity 29.2%; Pred. No. 1.3e-14;
Matches 57; Conservative 37; Mismatches 75; Indels 26; Gaps 6;

QY 107 WVQYIETPLLICDTKFDIROWFLVTDWNPITWIFKESYLRPSTQFSLDKL 163
Db 1 VWSKYIVDPLCIDGHKCDLRVYLVTSFDPDIYIYEEGIVRLATVYKIDRHADLNWPCW 60
QY 164 HLCNNAVQKYLKN-----DVGRSPLLPAHNMTSTTRFOEYLQKQ---RGAVWG- 209
Db 61 HLCNYSINKYHSDYIRSDAQDEDVG-----HKWTISALLRHLKLSQCDTRQLMNI 112

; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3621
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3621

Query Match 11.5%; Score 180; DB 2; Length 404;
Best Local Similarity 25.6%; Pred. No. 2.9e-11;
Matches 63; Conservative 33; Mismatches 84; Indels 66; Gaps 12;

QY 4 DTSADAVEDJTEAEWBDLTQYYSLVHGDAFISNSRYFSQCQALLNRITSVNPOT---D 60
Db 183 DAAAPALEDI---PW-----TSPGYLRPQVRL--RMEFFPETYRLD 219
QY 61 IDGLR-----NIWIIPAAKSRGRDIVCMRDVEITILELAA-----ADHPLSRDNK 105
Db 220 LKHEREAFFTLFDETQIWKPTASNOGKGFILRNQEEVAALQAKTRSMEDDPIHHTP 279
QY 106 W-----VQYIETPLLICDTKFDIROWFLVTDWNPITWIFKESYLRPSTQFSLDKL 159
Db 280 FRGQPARVQRYIQNPLLVNDRKFDVRSYLLIACITPYMI-FFGHGVARLTLSLYDPHSS 338
QY 160 DSAIHLCNNAVQKYLKNDVGRSP---LLPAHNMTSTTRFOEYL-----ORQGRGAVGMSV 211
Db 339 DLGGHLTNQPMQK-----KSPLYMLLKHTVMSMEHLNRYISDTFWKARGLAKDW--- 388
QY 212 IYPSMK 217
Db 389 VFTTLK 394

RESULT 5
US-09-270-767-62411
; Sequence 62411, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62411
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-62411
```


INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 626 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE:
US-08-596-300A-7

Query Match 6.0%; Score 94; DB 1; Length 626;
Best Local Similarity 20.1%; Pred. No. 0.24;
Matches 72; Conservative 40; Mismatches 126; Indels 120; Gaps 17;
QY 4 DTSADAVEDLTAEWEDLTQQYSLVHGDAFISNRNYSQCOA----- 47
DB 207 NTGFDLWEVNSSPFTIAASHRALVEGSAFAKSVGSSCSACDAPILCFQQSFWSNS 266
QY 48 --LLNRITSVNPTQDIDG--LRNIWIKPAKASGRDI-----VCMORVEBIELEAAADHPL 100
DB 267 GYIISFNVRSGKDINSVLTSIHNFDPAA---GCDVNTFQPCSDR-----ALANHKV 316
QY 101 SRDNK--W-----VVQKIETPLLICDTKFDIRQWFLVT-----DWNPLTIW 140
DB 317 VVDSMRFGVNSGRTAGKAAAVGRYAE-----DVYNGNPWYLATLAAAEQLYDAVYVW 370
QY 141 -----FYKESYLRFSTORFSLDKLDSAIHLCCNNAVOKY----- 173
DB 371 KKQGSITVTSTSLAFKDLVPSVSTGYTSSSSSTYTAIINAVTYADGFVDIVAQYTP 428
QY 174 -----LKNVGRSPLLPANHMTSTRFQYLRQGRGAV---WGSVIYPSMKKAIAH 222
DB 429 SDGSLAEQFDKDSG-APLSATHLTWSYASFLSAAARRA-GIVPPSWGAAASNSLPGSCS- 485
QY 223 AMKVAQDHVEPRKNSFELYGADFLGRDPRFWLIEINSPTHPSTPTVTAQLCAQVOE 280
DB 486 ASTVAGSATATATATSF-----PANLTPASTTVPPTQTGCAADHE 525

RESULT 13
US-08-596-300A-14
Sequence 14, Application US/08596300A
Patent No. 5834191
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Production of Heterologous Peptides
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/596.300A
FILING DATE: 13-FEB-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 1321-1-001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:
LENGTH: 626 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE:
US-08-596-300A-14

Query Match 6.0%; Score 94; DB 1; Length 626;
Best Local Similarity 20.1%; Pred. No. 0.24;
Matches 72; Conservative 40; Mismatches 126; Indels 120; Gaps 17;
QY 4 DTSADAVEDLTAEWEDLTQQYSLVHGDAFISNRNYSQCOA----- 47
DB 207 NTGFDLWEVNSSPFTIAASHRALVEGSAFAKSVGSSCSACDAPILCFQQSFWSNS 266
QY 48 --LLNRITSVNPTQDIDG--LRNIWIKPAKASGRDI-----VCMORVEBIELEAAADHPL 100
DB 267 GYIISFNVRSGKDINSVLTSIHNFDPAA---GCDVNTFQPCSDR-----ALANHKV 316
QY 101 SRDNK--W-----VVQKIETPLLICDTKFDIRQWFLVT-----DWNPLTIW 140
DB 317 VVDSMRFGVNSGRTAGKAAAVGRYAE-----DVYNGNPWYLATLAAAEQLYDAVYVW 370
QY 141 -----FYKESYLRFSTORFSLDKLDSAIHLCCNNAVOKY----- 173
DB 371 KKQGSITVTSTSLAFKDLVPSVSTGYTSSSSSTYTAIINAVTYADGFVDIVAQYTP 428
QY 174 -----LKNVGRSPLLPANHMTSTRFQYLRQGRGAV---WGSVIYPSMKKAIAH 222
DB 429 SDGSLAEQFDKDSG-APLSATHLTWSYASFLSAAARRA-GIVPPSWGAAASNSLPGSCS- 485
QY 223 AMKVAQDHVEPRKNSFELYGADFLGRDPRFWLIEINSPTHPSTPTVTAQLCAQVOE 280
DB 486 ASTVAGSATATATATSF-----PANLTPASTTVPPTQTGCAADHE 525

RESULT 14
US-09-248-796A-20939
Sequence 20939, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBIC
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248.796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 20939
LENGTH: 816
TYPE: PRT
ORGANISM: Candida albicans
US-09-248-796A-20939

Query Match 6.0%; Score 93.5; DB 2; Length 816;
Best Local Similarity 20.8%; Pred. No. 0.42;
Matches 64; Conservative 45; Mismatches 125; Indels 73; Gaps 13;
QY 8 DAVEDLTAEWEDLTQQYSLVHGDAFISNRNYSF----- 42
DB 348 DAFQDLLNSNTGDVV---IKASDGVFAHAFVLKARSAPFETLLSERWDTSEKGNVQYVD 404
QY 43 -----SQCQALLNRITSVNPTQDID-----GLRNIWIKPAKASGRDIVCMRVEEI 90
DB 405 FTGLTKFQVTLIRHLRYGVSNESLLDCFQYDFGSKOYFINDLBLEIEVADELLFQLKSV 464

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: April 4, 2006, 12:43:06 ; Search time 44.3857 Seconds
(without alignments)
4657.352 Million cell updates/sec
Title: US-10-635-977-2_COPY_73_365
Perfect score: 1560
Sequence: 1 EDIDTSADAVEDLTEAEWED.....LCAQVQEDTIKVAVDRSCDI 293

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2166443 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 2166443
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : UniProt 05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	1447	92.8	747	2	Q5J284	HUMAN	Q5J284 homo sapien
2	1303	83.5	518	2	Q8C0V2	MOUSE	Q8C0V2 mus musculus
3	1303	83.5	781	2	Q8CON7	MOUSE	Q8CON7 mus musculus
4	908	58.2	352	1	TTLI3	HUMAN	Q9V4I7 homo sapien
5	908	58.2	352	2	Q6AWA3	HUMAN	Q6AWA3 homo sapien
6	908	58.2	352	2	Q4KWS8	HUMAN	Q4KWS8 homo sapien
7	908	58.2	434	2	Q8NDN8	HUMAN	Q8NDN8 homo sapien
8	908	58.2	744	2	Q9H876	HUMAN	Q9H876 homo sapien
9	893	57.2	704	2	Q8BV51	MOUSE	Q8BV51 mus musculus
10	855.5	54.8	534	2	Q4RY08	TETNG	Q4RY08 tetraodon n
11	805	51.6	261	2	Q58CT2	BOVIN	Q58CT2 bos taurus
12	805	51.6	266	2	Q922T0	MOUSE	Q922T0 mus musculus
13	673	43.1	572	2	Q7Q156	ANOGA	Q7Q156 anopheles g
14	636	40.8	992	2	Q9VM91	DROME	Q9VM91 drosophila
15	539.5	34.6	501	2	Q7PMD3	ANOGA	Q7PMD3 anopheles g
16	515.5	33.0	719	2	Q9VM92	DROME	Q9VM92 drosophila
17	515.5	33.0	756	2	Q5BHY1	DROME	Q5BHY1 drosophila
18	439	28.1	281	2	Q6ZU95	HUMAN	Q6ZU95 homo sapien
19	413	26.5	331	2	Q5TNZ9	ANOGA	Q5TNZ9 anopheles g
20	385	24.7	101	2	Q96G88	HUMAN	Q96G88 homo sapien
21	344	22.1	523	2	Q54TUI	DICDI	Q54TUI dictyosteli
22	331	21.2	1075	2	Q7QTI3	GIALA	Q7QTI3 giardia lam
23	330	21.2	794	2	Q6BFH6	PARTI	Q6BFH6 paramacium
24	327.5	21.0	461	2	Q641W7	RAT	Q641W7 rattus norv
25	325	20.8	464	2	Q9D570	MOUSE	Q9D570 mus musculus
26	324	20.8	496	2	Q961I9	DROME	Q961I9 drosophila
27	324	20.8	496	2	Q9VX74	DROME	Q9VX74 drosophila
28	322	20.6	375	2	Q4V8C1	RAT	Q4V8C1 rattus norv
29	306.5	19.6	403	2	Q7QZC8	GIALA	Q7QZC8 giardia lam
30	299	19.2	1339	2	Q5F498	CHICK	Q5F498 gallus gall
31	297.5	19.1	827	2	Q9VKI9	DROSOPHILA	Q9VKI9 drosophila

ALIGNMENTS

RESULT 1

ID	Q5J284	HUMAN	PRELIMINARY;	PRT;	747	AA.	
AC	Q5J284						
DT	10-MAY-2005	(Tremblrel. 30, Created)					
DT	10-MAY-2005	(Tremblrel. 30, Last sequence update)					
DE	OTTHUMP0000028514	(Fragment)					
GN	ORFNames=RP3-355C18.2-002;						
OS	Homo sapiens (Human)						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;						
OC	Homo.						
OX	NCBI_TaxID=9606;						
RN	[1]						
RP	NUCLEOTIDE SEQUENCE.						
RA	Cobley V.;						
RL	Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.						
DR	EMBL; AL022327; CA142686.1; -; Genomic DNA.						
DR	GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.						
DR	GO; GO:0006464; P:protein modification; IEA.						
DR	InterPro; IPR004344; Tub_tyr_lygase.						
DR	Pfam; PF03133; TTL; 1.						
FT	NON_TER	1					
SQ	SEQUENCE	747	AA;	83854	MW;	F90948E159BBE589	CRC64;
Query Match							
Best Local Similarity 92.8%; Score 1447; DB 2; Length 747;							
Matches 281; Conservative 4; Mismatches 6; Indels 24; Gaps 4;							
QY	1	EDIDTSADAVEDLTEAEWEDLTQYYSLVH---GDAFI-----SNSRNYF	42				
Db	259	EDIDTSADAVEDLTEAEWEDLTQYYSLVQVPLGSSIVLCIFKIQKVMSPFPPTARD---	316				
QY	43	SCQALLNRITSVNPQTDIGLRNIWIIPKAASRGR---DIVCMRVVEEILELAADH	98				
Db	317	RQCQALLNRITSVNPQTDIGLRNIWIIPKAASRGRGSPDIVCMRVVEEILELAADH	376				
QY	99	PLSRDNKVVQKYLETPLLICDTKFDIPQWFLVTDNPLTIWFYKESYLRFSTQFSLDK	158				
Db	377	PLSRDNKVVQKYLETPLLICDTKFDIPQWFLVTDNPLTIWFYKESYLRFSTQFSLDK	436				
QY	159	LDSATHLCNNVQVLYKNDVGRSPILLPAHNNMTSTRFOYLQQRQGRGAVGWSVIYPSMKK	218				
Db	437	LDSATHLCNNVQVLYKNDVGRSPILLPAHNNMTSTRFOYLQQRQGRGAVGWSVIYPSMKK	496				
QY	219	ATAHAMKVAQDHVSPRKNSEFELGADFVLGRDPRFWLIEINSSPTMHPSTPVTQAQV	278				
Db	497	ATAHAMKVAQDHVSPRKNSEFELGADFVLGRDPRFWLIEINSSPTMHPSTPVTQAQV	556				
QY	279	QEDTIKVAVDRSCDI	293				
Db	557	QEDTIKVAVDRSCDI	571				

Q8IGW4 drosophila
Q8IPB2 drosophila
Q8C125 mus musculus
Q8CFV5 mus musculus
Q8CHB8 mus musculus
Q8PLV5 homo sapien
Q9UPZ4 homo sapien
Q6EMB2 homo sapien
Q6EEF3 cercopithec
Q5R978 pongo pygma
Q5VK47 homo sapien
Q8T417 drosophila
Q9VQV6 drosophila

Q8IGW4 DROME
Q8IPB2 DROME
Q8C125 MOUSE
Q8CFV5 MOUSE
Q8CHB8 MOUSE
Q9PLV5 HUMAN
Q9UPZ4 HUMAN
Q6EMB2 HUMAN
Q6EEF3 CERAE
Q5R978 PONPY
Q5VK47 HUMAN
Q8T417 DROME
Q960F9 DROME
Q9VQV6 DROME

32 297.5 19.1 828 2
33 297.5 19.1 1070 2
34 297 19.0 438 2
35 297 19.0 559 2
36 297 19.0 1333 2
37 296 19.0 1048 2
38 296 19.0 1226 2
39 296 19.0 1277 2
40 296 19.0 1295 2
41 296 19.0 1299 2
42 293 18.8 341 2
43 290.5 18.6 541 2
44 290.5 18.6 982 2
45 290.5 18.6 989 2

```

RESULT 2
Q8COV2 MOUSE
ID Q8COV2 MOUSE PRELIMINARY; PRT; 518 AA.
AC Q8COV2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Mus musculus adult male testis cDNA, RIKEN full-length enriched
DE library, clone:4930524K07 product:hypothetical tubulin-tyrosine ligase
DE containing protein, full insert sequence.
GN Name:1700019P01Rik;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleishmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Maehima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaiji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Suganara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,

```

```

RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Tota T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK029745; BAC26595.1; -; mRNA.
DR MGI; MGI:1922902; 170019P01Rik.
DR GO; GO:0016874; P.ligase activity; IEA.
DR GO; GO:0004835; F.tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P.protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_ligase.
DR Pfam; PF03133; TTL; 1.
DR Hypothetical protein; Ligase.
KW SEQUENCE 518 AA; 59319 MW; 540C045659FAC0D6 CRC64;
SQ
Query Match 83.5%; Score 1303; DB 2; Length 518;
Best Local Similarity 80.5%; Pred. No. 1.3e-102;
Matches 239; Conservative 29; Mismatches 25; Indels 4; Gaps 1;
QY 1 EDIDTSADAVEDLTEAEWEDLTQOYSLVHGDAFINSRNYFQCCALLNRITSVNPQTD 60
DB 129 EDIDVSEASTEALSEEWNDLTQOYLLVHGNAISITDSKSYFACQALLSKISSVNPQTE 188
QY 61 IDGLRNIIWIKPAKSRGRDIVCMRDRVEILELAAADHPLSRDNKWKVQKIETPLIICD 120
DB 189 IDGIRNIWIKPAKSRGRDIVCMRDRVENILSLVAADSQTTKDNKWKVQKIETPLIYD 248
QY 121 TKFDIQWFLVTDWNPDLTTFYKESYLRFSTQFSLDKLSATHLCNNVQKYLKNDVGR 180
DB 249 TKFDIRQWFLVTDWNPDLTTFYKESYLRFSTQFSLDKLSATHLCNNSIQRLKNDKER 308
QY 181 SPLLPAHNMWTSRFOYQLQRQGRGAVWGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 240
DB 309 SPLLPCHNMWTSRFOYQLQRGRGGTWSIIYPSMKRAVTNMRVAQDHVEARKNSFEL 368
QY 241 YGADFVLGRDFRPLWIEINSSPTMHPSTPVTALCAQVQSDTIKVAV----DRSCDI 293
DB 369 YGADFILGRDFKPLWIEINSSPTMHPSTPVTALCAQVQSDTIKVAVVVDVDRKLDNRNCDI 425
RESULT 3
Q8CON7 MOUSE
ID Q8CON7 MOUSE PRELIMINARY; PRT; 781 AA.
AC Q8CON7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Mus musculus adult male testis cDNA, RIKEN full-length enriched
DE library, clone:4933401B17 product:hypothetical tubulin-tyrosine ligase
DE containing protein, full insert sequence.
GN Name:1700019P01Rik;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

```



```
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kakakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RT Nat. Genet. 36:40-45(2004).
DR EMBL; AK023960; BAB14741.1; -; mRNA.
DR Ensembl; ENSG00000156983; Homo sapiens.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_lygase.
DR Pfam; PF03133; TTL; 1.
DR KW Ligase.
FT NON TER 744 744
SQ SEQUENCE 744 AA; 84683 MW; DF661753E4AFF0DF CRC64;
Query Match 58.2%; Score 908; DB 2; Length 744;
Best Local Similarity 56.7%; Pred. No. 1.2e-68;
Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;
QY 2 DIDSADVEDLLEAEWEDLTQYYSIVHGDAFISNSRNYFSQOALLNRITSVNPQTDI 61
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 DDKDLLEAPLYLTPGWSLEFLQRYQVVHGEALRHLDVTQVRCEDILQQLQAVVQIDM 273
QY 62 DGLRNIIWKPAKSRGRDVCMDRVEEILELAAADHPJLSRDNKVVQVYETPLICDT 121
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
274 EGRNIIWIKPGAKSRGIMCMCHLEEMKLVNGNPVVMKDGKVVVQVYERPLLIPT 333
QY 122 KFDIROWFLVDNPLTIWYKESYLRFSQFSLDKLSAHLCHNAVKYLNKDVGRS 181
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
334 KFDLRQWFLVDNPLTWVFRDSYIRFSTQSPFLSKNLDSVHLCHNSIQHLENSCHRH 393
QY 182 PLLPAHNMWTSRFFQELRQGRGAVNGSVIYPSMKKAI AHAMKVAQDHVPEKNSFELY 241
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
394 PLLPDDNMWSSQRFQAHLEKNGAPNAWSTIIVPGMKDAVHALQSDTVQCRKASFELY 453
QY 242 GADFVLGRDPRPWLIEINSSFTMHPSTPTVTAQLCAQVQEDTIKAVDRSCD 292
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
454 GADFVFGEDFPWLIEINASPTWAPSTAVTARLCAGVQADTLRAWVDRMLD 504
RESULT 9
Q8BV51 MOUSE
ID Q8BV51 MOUSE PRELIMINARY; PRT; 704 AA.
AC Q8BV51
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
DE library, clone:A630053H17 product:HOTT1 PROTEIN homolog.
GN Name=A63344J24Rik;
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RN NUCLEOTIDE SEQUENCE.
```

```
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Iwaza M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fletschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakomoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyoko-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RN Nature 409:685-690(2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Yumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Matsumoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa K., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK080321; BAC37878.1; -; mRNA.
DR Ensembl; ENSMUSG00000030276; Mus musculus.
DR MGI; MGI:2141418; 483344J24Rik.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
```


QY 180 RSPLLPAHNNWTSRFOEYLQROGRGAVMGSVIYPSMKKAIAHAMKVAQDHVEPRKNSPE 239
DB 122 RHPLLPDDNWSSQKQAHQETGAPNWSVIVPGMKAAVIAHALQTSQDTVOCRKASPE 181
QY 240 LYGADFVGLGRDPRFWLIEINSSPTMHPSTPVTALCAQVQEDTIKVAVDKSCD 292
DB 182 LYGADFVFGEDFQFWLIEINASTPMTAPSTAVTARLCAGVQADTLRVVIDWRD 234

RESULT 12
Q922T0 MOUSE
ID Q922T0 MOUSE PRELIMINARY; PRT; 266 AA.
AC Q922T0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 4833441J24Rik protein.
DE 4833441J24Rik;
GN Names=4833441J24Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzyzinski M.I., Skalek U., Smailus D.E.,
RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RG NIH MGC Project;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006830; AAH06830.1; -; mRNA.
DR Ensembl; ENSMUSG0000030276; Mus musculus.
DR MGI; MGI:2141418; 4833441J24Rik.
DR GO; GO:0004835; F:ubulin-cytosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_ligase.
DR Pfam; PF03133; TTL; 1.
SQ SEQUENCE 266 AA; 30506 MW; F8E8FB52FA8B8E98 CRC64;

Query Match 51.6%; Score 805; DB 2; Length 266;
Best Local Similarity 61.0%; Pred. No. 2.1e-60;
Matches 144; Conservative 41; Mismatches 47; Indels 4; Gaps 1;

QY 61 IDGLRNWIITKPAKSRGRDIVCMRVEIILELAADHPLSRDNKVVQKVIETPLICD 120
DB 1 MEGQNIWIIVKPAKSRGRDIVCMRVEIILELAADHPLSRDNKVVQKVIETPLIFG 60
QY 121 TKFDIRQWFLVTDWNPITWYKESYLRFSTQRFSLDKLSAHLCCNNAVQKYLKNDVGR 180
DB 61 TKFDLRQWFLVTDWNPITWYKESYLRFSTQRFSLDKLSAHLCCNNAVQKYLKNDVGR 120

QY 181 SPILLPAHNNWTSRFOEYLQROGRGAVMGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 240
DB 121 HPMPLPDDNWSSORFOAHLEQVDAPKAWSSVIVPGMKAAVIAHALQTSQDNVQCRKASFEL 180
QY 241 YGADFVGLGRDPRFWLIEINSSPTMHPSTPVTALCAQVQEDTIKVAV----DRSCD 292
DB 181 YGADFVFGEDFQFWLIEINASTPMTAPSTAVTARLCAGVQADTLRVVIDRDLRSCD 236

RESULT 13
Q7Q156 ANOGA
ID Q7Q156 ANOGA PRELIMINARY; PRT; 572 AA.
AC Q7Q156;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ENSANGP0000022337 (Fragment).
DE ORFNames=ENSANG0000019848;
GN Anopheles gambiae str. PEST.
OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
OC Anophelinae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RT "Anopheles gambiae re-annotation."
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.
DR EMBL; ABA01008980; EAA13905.2; -; Genomic DNA.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_ligase.
DR Pfam; PF03133; TTL; 1.
FT NON_TER 1 1
FT NON_TER 572 572
SQ SEQUENCE 572 AA; 67109 MW; 8938D9DD5935071 CRC64;

Query Match 43.1%; Score 673; DB 2; Length 572;
Best Local Similarity 45.0%; Pred. No. 1.1e-48;
Matches 135; Conservative 54; Mismatches 97; Indels 14; Gaps 5;

QY 2 DIDTSADAVEDLTEAEWEDLTQYYISLVHGD---AFISNSR-----NYFSOQALLNRI 52
DB 239 DIDIEEDT--KVDHDDWDFVLTTHYLLTHEDNRQLLKEEREADAIEHYLAESAQSLEQI 296
QY 53 TSNVPQTDIDGLRNWIITKPAKSRGRDIVCMRVEIILELAADHPLSRDNKVVQKVI 112
DB 297 KSWHPQVALDGLNINIVKSGNCRGHIHLMNNIKI--IAMVNPVSKTRVIQKVI 354
QY 113 ETPLLICDITKFDIRQWFLVTDWNPITWYKESYLRFSTQRFSLDKLSAHLCCNNAVQK 172
DB 355 ERPLIHTKFDIRQWFLVTDWNPITWYKESYLRFSTQRFSLDKLSAHLCCNNAVQK 414
QY 173 YLKNVDCRSPLIPAHNNWTSRFOEYLQROGRGAVMGSVIYPSMKKAIAHAMKVAQDHVE 232
DB 415 KYHNAV-RDRLPHENNMWDCHTFQAYLRQIDKVMWSEIYPGMKAIIIGSLACQDNMD 473
QY 233 PRKNSFELYGADFVGLGRDPRFWLIEINSSPTMHPSTPVTALCAQVQEDTIKVAVDKSCD 292
DB 474 RRPNTFELYGADFVGLGRDPRFWLIEINSSPTMHPSTPVTALCAQVQEDTIKVAVDKSCD 533

RESULT 14
Q9VM91_DROME

ID Q9VM91_DROME PRELIMINARY; PRT; 992 AA.
 AC Q9VM91;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CG11323-PA.
 GN Name=CG11323; ORFNames=CG11323;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Brill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burdick K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft A., Kravitz S., Kulp D., Lal Z.,
 RA Lasko F., Lei Y., Levitsky A.C., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M.C., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195 (2000).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodson A.,
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirskaas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*
 RT melanogaster euchromatic genome sequence.";
 RL Genome Biol. 3:RESEARCH0073-RESEARCH0079 (2002).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskaas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.;
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
 RT a genomics perspective."

RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20 (2002).
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).
 RN [5]
 RP NUCLEOTIDE SEQUENCE.
 RG Berkeley *Drosophila* Genome Project;
 RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
 RA Hoskins R., Stapleton M., Pacle J., Park S., Svirskaas R., Smith E.,
 RA Yu C., Rubin G.;
 RT "Drosophila melanogaster release 4 sequence.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP NUCLEOTIDE SEQUENCE.
 RG FlyBase;
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003614; AAF52432.1; -; Genomic DNA.
 DR Ensembl; CG11323; *Drosophila melanogaster*.
 DR FlyBase; FBgn0031854; CG11323.
 DR GO; GO:0004835; F: tubulin-tyrosine ligase activity; IEA.
 DR GO; GO:0006464; P: protein modification; IEA.
 DR InterPro; IPR004344; Tub_tyr_lygase.
 DR Pfam; PF03133; TTL; 1.
 SQ SEQUENCE 992 AA; 112255 MW; 0FC34674C15BDAAD CRC64;
 Query Match 40.8%; Score 636; DB 2; Length 992;
 Best Local Similarity 42.8%; Pred. No. 3e-45;
 Matches 128; Conservative 59; Mismatches 96; Indels 16; Gaps 5;
 QY 2 DDTSDAIVED--LTAEWEDLTQYSLVHGDAFISN-----RNYSQCALNRT 53
 DB 349 DID-----FEDPPKIWEHDWDAFLFQHOOLVNEGDRIQHOGGQRLEPVMKSLSLVDKMK 403
 QY 54 SVNPQTIDGLRNIIWIKPAKSRGDIIVCMRDVEEILELAAADHPLSRDNKWKVQKXIE 113
 DB 404 VHMQYSLDGYQNMWIKPANKRGRIILMDLKKI--LGVNLSIASKSRVYVQKXIE 461
 QY 114 TPLLICDTKFDIROWFLVTDNPLTIWPKSYLRFSTQFSLDKLDSAIHLCNNVQKY 173
 DB 462 RPLILFQTKFDIROWFLTIINTQPLVWFWFYRESYLRFSSQYSLSNHHSVHLTNYAIQKK 521
 QY 174 LKNDVGRSPLPAHNMWTSITFQYLRQGRGAVGWSVIYPSMKKAIHAHMKVAQDVPEP 233
 DB 522 YTNQ-KDKRLPSPNMMDVCSFYQALRQIGKNMMLERIFFPMRKAIYVGCMLASQENMDR 580
 QY 234 RKNSELYGADFVLGRDFPMLIENGSPTHPTVTAOLCAOVQEDTIKVAVDKSCD 292
 DB 581 RPNTFELFGADFMCENFYPLIENGSPDLGATTSVTARMCPQCLVEDVVKVVIDRRTD 639
 RESULT 15
 ID Q7PMD3 ANOGA
 AC Q7PMD3;
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DE ENSANGP0000022334 (Fragment).
 GN ORFNames=ENSANGG00000019845;
 OS Anopheles gambiae str. PEST.
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Culicidae;
 OC Anophelinae; Anopheles.

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:45:31 ; Search time 10.8922 Seconds
(without alignments)
2588.231 Million cell updates/sec

Title: US-10-635-977-2_COPY_73_365
Perfect score: 1560
Sequence: 1 EDIDTSADAVEDLTEAWED.....LCAQVEDTIKVAVDRCDI 293

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	908	58.2	352	2 T12515	hypothetical prote
2	255.5	16.4	379	2 A45443	tubulin-tyrosine 1
3	244	15.6	640	2 E89575	protein ZK1128.6 {
4	244	15.6	680	2 T27699	hypothetical prote
5	209.5	13.4	1198	2 T20262	hypothetical prote
6	209.5	13.4	1203	2 C9217	protein C55A6.2 {i
7	166	10.6	662	2 T20343	hypothetical prote
8	160.5	10.3	403	2 T37571	tubulin-tyrosine 1
9	155.5	10.0	883	2 A96805	hypothetical prote
10	144.5	9.3	753	2 S48261	hypothetical prote
11	95.5	6.1	2176	2 T39188	probable U5 snRNP-
12	95	6.1	1217	2 T21403	hypothetical prote
13	94.5	6.1	1528	2 S13743	DNA strand transfe
14	94	6.0	626	2 T49625	glucan 1,4-alpha-g
15	94	6.0	626	2 S36364	glucan 1,4-alpha-g
16	93	6.0	457	2 T29116	hypothetical prote
17	91.5	5.9	4660	2 T42737	gp330 protein prec
18	90.5	5.8	816	2 T19049	hypothetical prote
19	89	5.7	2870	2 H96974	cyclic beta 1-2 gl
20	88	5.6	282	2 E84317	hypothetical prote
21	87.5	5.6	1270	2 T22615	hypothetical prote
22	87.5	5.6	3971	2 T44806	myosubtilin synth
23	87.5	5.6	4688	2 F82885	hypothetical prote
24	87	5.6	733	2 E71937	flagellar biosynth
25	87	5.6	1044	2 S16516	integrin alpha-8 c
26	86.5	5.5	733	2 A64650	flagellar biosynth
27	86.5	5.5	796	2 T16367	hypothetical prote
28	86.5	5.5	1857	1 S01787	fatty-acid synthas
29	86	5.5	295	2 T30703	probable mRNA guan

30	85.5	5.5	1124	2 JH0588	calmodulin-binding
31	85	5.4	590	2 H70130	oligoendopeptidase
32	84.5	5.4	445	2 I46225	mucin A, tracheal
33	84.5	5.4	522	2 T45824	hypothetical prote
34	84	5.4	439	2 T01270	hypothetical prote
35	84	5.4	1872	2 T30888	vitellogenin - Ath
36	84	5.4	2182	2 T14320	calcineurin inhibi
37	83.5	5.4	930	2 A25923	progesterone recep
38	83	5.3	4128	2 JC6306	protein kinase (EC
39	82.5	5.3	275	2 G75130	translation initia
40	82.5	5.3	349	2 F91218	probable transport
41	82.5	5.3	349	2 H86064	probable transport
42	82.5	5.3	349	2 D65182	probable transport
43	82.5	5.3	396	2 T04561	hypothetical prote
44	82.5	5.3	609	2 B84783	probable poly(A) b
45	82.5	5.3	1297	2 T52065	probable myb-relat

ALIGNMENTS

RESULT 1

T12515
hypothetical protein DKFp434B103.1 - human
C:Species: Homo sapiens (man)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 09-Jul-2004
C:Accession: T12515
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gasteinhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z17524
A:Accession: T12515
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-352 <WAM>
A:Cross-references: UNIPROT:O9Y4R7; UNIPARC:UPI00001377A6; EMBL:AL096725
A:Experimental source: adult testis; clone DKFp434B103
C:Genetics:
A:Note: DKFp434B103.1

Query Match 58.2%; Score 908; DB 2; Length 352;
Best Local Similarity 56.7%; Pred: No. 7.2e-72;
Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;

QY	2	DIDTSADAVEDLTEAWEDLTQQYSLVHGDAFISNRNFSQCQALLNRITSVNPQTDI	61
Db	2	DIDKDLAPLYLTPEGWSLFLQRYQVHVHGALRHLDTVQRCEDIQLQLAQVVPQIDM	61
QY	62	DGLRNIIKPAAKSRGRDIVCMORVEIILEAAADHPLSRDNKRWVQKYIETPLLICDT	121
Db	62	EGDRNIWIKFGAKSRGRGIMCMDHLEMLKLVNGNPNVVMKDKWVQKYIERPLLIQGT	121
QY	122	KFDIRQWFLVTDWNPPLTTFYKESYLRESTQSFSLDKLDSAIHLCCNNAVOKYLKNDVGRS	181
Db	122	KFDLRQWFLVTDWNPPLTTFVYRDSYIFRESTQFSLKLNLDNSVHLCCNNSIQHLENSCHRH	181
QY	182	PLLPANHWMTSTRFOYLQROGRGAVMGSVTPYSMKKAIAHAMKVQADHVEPRKNSFELY	241
Db	182	PLLPDPNWSQORQAHLENGAPNANSTIIIVPGMKDAVIALQTSQDTVQCRKASFELY	241
QY	242	GADFVLGRDFPWLIEINSSPTMHPSTPTVTAQLCAQVEDTIKVAVDRCSD	292
Db	242	GADFVFGEDFPWLIEINASPTMAPSTAVTARLCAGVQADTLRVVIDRMLD	292

RESULT 2

A45443
tubulin-tyrosine ligase (EC 6.3.2.25) [validated] - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A45443
R:Ersfeld, K.; Wehland, J.; Plessmann, U.; Dodemont, H.; Gerke, V.; Weber, K.
J. Cell Biol. 120, 725-732, 1993
A:Title: Characterization of the tubulin-tyrosine ligase.

A:Reference number: A45443; MUID:93147125; PMID:8093886
A:Accession: A45443
A>Status: preliminary
A:Molecule type: mRNA; protein
A:Residues: 1-379 <ERS>
A:Cross-references: UNIPROT:P38160; UNIPARC:UPI00001377A9; GB:X68453; GB:SS4050; NID:g21
A:Experimental source: brain
A>Note: sequence extracted from NCBI backbone (NCBIN:123858, NCBIF:123859)
C:Keywords: ligase

Query Match 16.4%; Score 255.5; DB 2; Length 379;
Best Local Similarity 28.4%; Pred. No. 1.3e-14;
Matches 74; Conservative 46; Mismatches 98; Indels 43; Gaps 11;

QY 36 SNSRNFSCQALLNRTISVPTDIDGLRNIIWIKPAAKSRGRDIVCMRDVEEILELAA 95
DB 125 TTEREF-----LTSYNKKE-DGEGNWIAKSSAGAKGEGILISEATELDDFI- 173
QY 96 ADHPLSRDNK---WVVKYIETPLLI--CDTKFDIQWFLVTDWNPITTFWFKESYLRS 150
DB 174 -----DNQGVHVIOKYLRLLELLEPGHRKFDIRSVLVD--HOYNIYLYREGVLRTA 224
QY 151 TQRFSLDKI-DSAIHLCLNNAVQKYLKNDVGRSPLLPANNNMTSTRFQEVYLQRCRGAVWG 209
DB 225 SEPYHTDNFQDKTCHLTNHCIOKEYSKNYGK---YEGNEMFFEEFNQYLT-----SALN 276
QY 210 SVIYPSMKKAIAHAMVAQDHVPR-----KNSFELYGADPVLGRDPRPMLIEINSSP 262
DB 277 ITUESILLIQIKHIIKRSCLLSVEPAISTRLHPYQSOLFQGFDPVDEDLKWLLIEVNGAP 336
QY 263 TMHPSTPVTQAQVQEDTI 283
DB 337 AC--AQKLYAELCOGIVDIAI 355

RESULT 3
E88575
protein ZK1128.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: E88575
C:Anonymous, The C. elegans Sequencing Consortium.
R:Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:99069613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.elegans/
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: E88575
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-640 <STO>
A:Cross-references: UNIPROT:Q09647; UNIPARC:UPI000017A077; GB:chr_III; PIDN:CAA87425.1;
C:Genetics:
A:Gene: ZK1128.6
A:Map position: 3

Query Match 15.6%; Score 244; DB 2; Length 640;
Best Local Similarity 27.6%; Pred. No. 2.7e-13;
Matches 67; Conservative 48; Mismatches 90; Indels 38; Gaps 9;

QY 60 DIDGLRNIIWIKPAAKSRGRDIVCMRDVEEILELAAADHPLSRDNKVVQKIETPLLIC 119
DB 283 ETDASRHV-IVKPPASARGTGISVTRKPKDPTTATL-----VAQHYIERPLTIN 331
QY 120 DTKFDIQWFLVTDWNPITTFWFKESYLRSFSTQRFSLDK---LDSAIHLCLNNAVQKYLKN 176
DB 332 RAKFDLRLYAYVPTFEPLRVIIYDQGLVRFASVPYSHSVSTISNKYMHLLTNYINKLAEA 391
QY 177 D-VGRSPL--LPAHNNMTSTRF-----OEYLQRCRGAVWGSVIYPSMKKAIAHAMKVA 227
DB 392 DGVANKVPKWTLLHLW--EHFDEMGVDRKIQRE-----IEEVIKAFISTEKPI 440
QY 228 QDH-----VEPRKNSFELYGADPVLGRDPRPMLIEINSSPTMHPSTPVTQAQVQEDTI 283
DB 481 REHMSRFLQEFCYELFGIDIILDEDYKWPALLEVNISPSLSHSGTPLDVSVKAPLAKDVL 540

RESULT 4
T20262
hypothetical protein ZK1128.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T20262
R:Kershaw, J.
submitted to the EMBL Data Library, January 1995
A:Reference number: Z20407
A:Accession: T20262
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-680 <WIL>
A:Cross-references: UNIPROT:Q09647; UNIPARC:UPI000017A076; EMBL:Z47357; PIDN:CAA87425.2
A:Experimental source: clone ZK1128
C:Genetics:
A:Gene: CBSP:ZK1128.6
A:Map position: 3
A:Introns: 114/1; 208/3; 242/3; 334/3; 362/2; 405/2; 424/2; 550/2; 633/2

Query Match 15.6%; Score 244; DB 2; Length 680;
Best Local Similarity 27.6%; Pred. No. 2.9e-13;
Matches 67; Conservative 48; Mismatches 90; Indels 38; Gaps 9;

QY 60 DIDGLRNIIWIKPAAKSRGRDIVCMRDVEEILELAAADHPLSRDNKVVQKIETPLLIC 119
DB 323 ETDASRHV-IVKPPASARGTGISVTRKPKDPTTATL-----VAQHYIERPLTIN 371
QY 120 DTKFDIQWFLVTDWNPITTFWFKESYLRSFSTQRFSLDK---LDSAIHLCLNNAVQKYLKN 176
DB 372 RAKFDLRLYAYVPTFEPLRVIIYDQGLVRFASVPYSHSVSTISNKYMHLLTNYINKLAEA 431
QY 177 D-VGRSPL--LPAHNNMTSTRF-----OEYLQRCRGAVWGSVIYPSMKKAIAHAMKVA 227
DB 432 DGVANKVPKWTLLHLW--EHFDEMGVDRKIQRE-----IEEVIKAFISTEKPI 480
QY 228 QDH-----VEPRKNSFELYGADPVLGRDPRPMLIEINSSPTMHPSTPVTQAQVQEDTI 283
DB 481 REHMSRFLQEFCYELFGIDIILDEDYKWPALLEVNISPSLSHSGTPLDVSVKAPLAKDVL 540

QY 284 KVA 286
DB 541 NLA 543

RESULT 5
T20262
hypothetical protein C55A6.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T20262
R:Kershaw, J.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19243
A:Accession: T20262
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1198 <WIL>
A:Cross-references: UNIPROT:O17720; UNIPARC:UPI000017A066; EMBL:Z81051; PIDN:CAB02862.2
A:Experimental source: clone C55A6
C:Genetics:
A:Gene: CBSP:C55A6.2
A:Map position: 5
A:Introns: 32/3; 68/1; 107/3; 201/3; 273/1; 364/3; 407/3; 460/3; 476/2; 534/2; 677/3; 7

Db 149 SMCDAQGLRSTIEELQAFDSFDESESEAGLEBKGDITVAFNNKIVISQIRNFL 208
QY 108 VQKYIETPLLCIDTKFDIRQWFLVTDNPLTIWFKESYLRFSQRFSLDKLDSAI---H 164
Db 209 VQKYISKPLLDHRKPHIRAYVLAT--GALSVLFFNEMCLLARDKYKPTDPDLFSH 266
QY 165 LCNNAVQYKLVNDVGRSPLLPAHNMWTSRFBQYLRQGRGAVGWGVIYPSMKKAIAHAM 224
Db 267 LSNATCLQ---GDNVEQSSI---RDFWNTS-----IENK-----DDIFKSLNIIIGDVF 308
QY 225 KVAQD---HVEPRKNSFELYGADFVLGRDPRFLWLEINSPPTMHPSTPVTALCAQVOE 280
Db 309 EAAATQGIHFQPLENCFELGVDFLVDCESQVILLEVNS----YPDFKQTGKLSNIIIE 364
QY 281 DTIKVAVD 288
Db 365 NUFSAVVE 372
RESULT 9
A96805
hypothetical protein TSM16.14 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: A96805
R:Theologias, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A96805
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-883 <STO>
A:Cross-references: UNIPROT:Q9CAQ1; UNIPARC:UPI00000A0A9A3; GB:AB005173; NID:96382502; PT
C:Genetics:
A:Gene: TSM16.14
A:Map position: 1
Query Match 10.0%; Score 155.5; DB 2; Length 883;
Best Local Similarity 25.3%; Pred. No. 2.4e-05;
Matches 55; Conservative 36; Mismatches 93; Indels 33; Gaps 7;

QY 62 DGLRNWIIPKAASGRDIVCMDRVBEILELAAADHPLSRDNKVVQKYIETPLLCIDT 121
Db 641 DQLNNLWILKPNMARTIDTSITDLSAIIR-----MMETGPKICQKYIEHPALPKGN 693
QY 122 KFDIRQWFLVTDNPLTIWFKESYLRFSQRFSLDKLDSAIHLGNNAVQYKLVNDVGRS 181
Db 694 KFDLRYVVLVRSIDPLIEIYLFIEFWRLSNPNYSLEKHSFFETHFTVMNY-----GRK 748
QY 182 PLLPAHNMWTSRFBQYLRQSG-----RGAVWGSV-----IYPSMK---KAIAHAMK 225
Db 749 ----LNHKPTAEFRFEQEHNDFYAFHFNNTQLSIVKWMIDHEKVQVIRAVFEAAA 804
QY 226 VAQDHVEPRKNSFELYGADFVLGRDPRFLWLEINSSP 262
Db 805 LAHPMQSPK-SRANYGVDVWLDSSFPFKILEVTYCP 840

RESULT 10
S48261
hypothetical protein YBR094w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YBR0821
C:Species: Saccharomyces cerevisiae
C:Date: 01-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 09-Jul-2004

C:Accession: S48261; S45962; S41800; S44676
R:Mannhaupt, G.; Stucka, R.; Ehnlle, S.; Vetter, I.; Feldmann, H.
Yeast 10, 1363-1381, 1994
A:Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
A:Reference number: S48255; MUID:95208357; PMID:7900426
A:Accession: S48261
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-753 <MAN>
A:Cross-references: UNIPROT:P38254; UNIPARC:UPI000013A3DD; EMBL:X78993; NID:9476045; PI
R:Feldmann, H.; Mannhaupt, G.; Schwarze, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45927
A:Accession: S45962
A:Molecule type: DNA
A:Residues: 1-753 <PE2>
A:Cross-references: UNIPARC:UPI000013A3DD; EMBL:Z35963; NID:9536366; PIDN:CAA85047.1; PID
R:Dekker, P.J.T.; Hoekert, W.; van Oosterum, K.; Grivell, L.A.
submitted to the EMBL Data Library, December 1992
A:Reference number: S31313
A:Accession: S41800
A:Molecule type: DNA
A:Residues: 167-351, 'TPE', 356-449, 'A', 451-562, 'R', 564-753 <DEK>
A:Cross-references: UNIPARC:UPI0000168D37; EMBL:X69881; NID:94086; PIDN:CAA49508.1; PID
C:Genetics:
A:Cross-references: SGD:S0000298
A:Map position: 2R
Query Match 9.3%; Score 144.5; DB 2; Length 753;
Best Local Similarity 21.8%; Pred. No. 0.00018;
Matches 57; Conservative 44; Mismatches 105; Indels 56; Gaps 9;
QY 68 WIIPKAASGRDIVCMDRVET-----LELAAADHP----- 99
Db 476 WIVKPSMSDKQGIRVFKFTIEDLQAFDSFDESEAEESGNDADDVNGEFMDNNKVN 535
QY 100 LSRDNKVVQKYIETPLLCIDTKFDIRQWFLVTDNPLTIWFKESYLRFSQRF-- 154
Db 536 ISQIRHFIIQEVLTNPELLASMDNRHFIICY--VVCGRDLQVFFVYDRMLALFAAKFPVP 593
QY 155 -----SLDKLDSAIHLGNNAVQYKLVNDVGRSPLLPAHNMWTSRFP--EYLQGRG 205
Db 594 LDPYAVSVTDLKLECHLTNTCLQS-KKKDKSSVL-----EFDSEIEIPNERKS 642
QY 206 AVWGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFELYGADFVLGRDPRFLWLEINSSPTMH 265
Db 643 NI-KEQIHSTINDVFLAAVNVNRLNFQPLPNAFETYGVDFLDSNVEVKKLEINAFDFK 701
QY 266 PSTPVTAAQLCAQVOEDTIKVAV 287
Db 702 QTGKDLKNLIDELFDDTVKYCV 723
RESULT 11
T39188
Probable U5 snRNP-specific 200kd protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T39188
R:Wedler, H.; Dueterhose, A.; Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1999
A:Reference number: Z21834
A:Accession: T39188
A>Status: preliminary; translated from GB/EMBL/DDDBJ
A:Molecule type: DNA
A:Residues: 1-2176 <WED>
A:Cross-references: UNIPROT:Q9UT24; UNIPARC:UPI000006BCA7; EMBL:AL121764; PIDN:CAB57421
C:Genetics:
A:Gene: SPDB:SPAC9.03c
A:Map position: 1
Query Match 6.1%; Score 95.5; DB 2; Length 2176;

Best Local Similarity 19.1%; Pred. No. 15;		Matches 65; Conservative 55; Mismatches 120; Indels 101; Gaps 15;	
QY	2	DIDTSADAVEDLTE---AEWEDLT-----QQYSLVHGDAFISNRYFSQCOALIN	50
Db	236	EADSEEAEEVEEDVLETSISQEEKKNENPDTEVI-----FISADTKKVT	291
QY	51	RITSVNPQTDIGLRNIWIIPKAASGRGDIQVMDRVEILELAADHPLSRDNKVVQK	110
Db	292	EIPTHVR-EIDA---FWIQREIAKYFADAVVVCQEKTNQAFALSADYDLGS-----	339
QY	111	YIETPLLICDTKFDIROWFLV---TDWNPLTIWFKY-----ESVLRSTQRF	154
Db	340	-LENELM---SIFDVEHFVLQVLLTKNRWTIVSCWMLKRAATDEERGLGVEEQIRAGRSW	395
QY	155	SLDKL-----DSAIHLCLNNAVQKYLKNDVGRSPL-----LPAHNMWTSRFRQ	196
Db	396	ILEALRPGAITIPDGLNELNNVVEKABPAPVSEIPLSKTLTSHKIVPKHQV-----DLR	451
QY	197	EYLQQRGAGVNGSVIYPSMKKAIAMKVAQDHVPRKNSPELYGADP-----	245
Db	452	NYVFTEG-----SRLMSNKAVKLPEGSPRRTCKGYEETHVPAPNKA	492
QY	246	VLGRDFRPLWLEINSSPTMHPSTPVTQAQLCAQVQEDTIKVA	286
Db	493	VLGADER--LVKIKELPEWSHOAFLNTQSLNRIQSHLYPIA	531
RESULT 12			
T21403			
hypothetical protein F26D2.10 - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004			
C:Accession: T21403			
R:McMurray, A.			
submitted to the EMBL Data Library, November 1996			
A:Reference number: Z19418			
A:Accession: T21403			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-1217 <WIL>			
A:Cross-references: UNIPROT:Q9XV62; UNIPARC:UPI0000060FE4; EMBL:Z81513; PIDN:CAN04180.1;			
A:Experimental source: clone F26D2			
C:Genetics:			
A:Gene: CESP:F26D2.10			
A:Map position: 5			
A:Introns: 29/1; 568/2; 608/3; 746/1; 871/3; 910/1; 985/1; 1018/3; 1124/1; 1170/3			
A:Superfamily: Caenorhabditis elegans hypothetical protein F36D3.5			
Query Match			
Best Local Similarity 6.1%; Score 95; DB 2; Length 1217;			
Matches 58; Conservative 49; Mismatches 106; Indels 68; Gaps 12;			
QY	3	IDTSADAVEDLTEAEWEDLTQQYSLVHGDAFISNRYN-----FYSQ	44
Db	78	LNTVVNVNVVASELSRMRHQEYTKLIH---FFENRTOSSFNDDVNAIFNNIIFILYFSR	134
QY	45	COALINRITSVNPQTDIGLRNIWIIPKAASGRGDIQV---MDRVEILELAADH--PL	100
Db	135	KSSLLSEVSSISTKNLNETKSI-----AKDLFNSCSLLTGLDFVALFDDDLQTPN	188
QY	101	SRDNKVVQKYIETPLLICD-----TKPD-----IROWFLVTDWNPLTIWFKYSYLRF	150
Db	189	SEKIAQIINEPHRQASLMTECTISGLKLDDEDIDALSNNKVNVEAYNKL--WVVKDSIILFE	246
QY	151	TQRLSLDKLDSAIHLCLNNAVQKYLKNDVGRSPLPAHNMWTSRFRQVQLQGRGAVNGS	210
Db	247	TFREPLGS-----WNESLLKIL-----SAFOTAKPLMNS-----RSSLGQS	282
QY	211	VIYPSMKKAIAMKVAQDHVPRKNSPELYGADPVLGRDF	251
Db	283	KIFEMAQ-LNKSIVEAEDHNISSINQLDKTKTAGFKNPGDF	322

RESULT 13

S13743

DNA strand transferase 2 - yeast (Saccharomyces cerevisiae)
N:Alternate names: DST2 protein; KEM1 protein; protein G1645; protein YGL173c; PAR5 pr
C:Species: Saccharomyces cerevisiae
C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004
C:Accession: S13743; S16701; S16885; S242156; S52457; A39790; S59238; S64190
R:Kim, J.; Ljungdahl, P.O.; Fink, G.R.
Genetics 126, 799-812, 1990
A:Title: kem mutations affect nuclear fusion in Saccharomyces cerevisiae.
A:Reference number: S13743; MUID:91169260; PMID:2076815
A:Accession: S13743
A:Molecule type: DNA
A:Residues: 1-1528 <KIM>
A:Cross-references: UNIPROT:P22147; UNIPARC:UPI000012DD97; EMBL:X54717; NID:g3846; PID
R:Tiehkoff, D.X.; Johnson, A.W.; Kolodner, R.D.
Mol. Cell. Biol. 11, 2593-2608, 1991
A:Title: Molecular and genetic analysis of the gene encoding the Saccharomyces cerevis
A:Reference number: S16701; MUID:91203880; PMID:1840632
A:Accession: S16701
A:Molecule type: DNA
A:Residues: 1-1528 <MOL>
A:Cross-references: UNIPARC:UPI000012DD97; GB:M58367; NID:g172579; PIDN:AAA35036.1; PI
R:Kipling, D.; Tambini, C.; Kearsey, S.E.
Nucleic Acids Res. 19, 1385-1391, 1991
A:Title: Rar mutations which increase artificial chromosome stability in Saccharomyces
A:Reference number: S16885; MUID:91227124; PMID:2027746
A:Accession: S16885
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1528 <KIP>
A:Cross-references: UNIPARC:UPI000012DD97; EMBL:X61181; NID:g4412; PIDN:CAA43487.1; PI
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1991
R:Larimer, F.W.; Stevens, A.
submitted to the EMBL Data Library, March 1992
A:Description: Structure of the XRN1 exoribonuclease gene of Saccharomyces cerevisiae.
A:Reference number: S42156
A:Accession: S42156
A:Molecule type: DNA
A:Residues: 1-1528 <LAR>
A:Cross-references: UNIPARC:UPI000012DD97; EMBL:M90097; NID:g173191; PIDN:AAA35219.1;
R:Bertani, I.; Coglievina, M.; Zaccaria, P.; Klima, R.; Bruschi, C.V.
submitted to the EMBL Data Library, February 1995
A:Description: The sequence of 11.1kb fragment on the left arm to Saccharomyces cerevi
A:Reference number: S52454
A:Accession: S52457
A:Molecule type: DNA
A:Residues: 1-1528 <BER>
A:Cross-references: UNIPARC:UPI000012DD97; EMBL:X84705; NID:g577853; PIDN:CAA59180.1;
R:Dykstra, C.C.; Kitada, K.; Clark, A.B.; Hamatake, R.K.; Sugino, A.
Mol. Cell. Biol. 11, 2583-2592, 1991
A:Title: Cloning and characterization of DST2, the gene for DNA strand transfer protei
A:Reference number: A39790; MUID:91203879; PMID:1850100
A:Accession: A39790
A:Molecule type: DNA
A:Residues: 1-1260, 1', 1262-1528 <DYK>
A:Cross-references: UNIPARC:UPI0000178F97; GB:M36725
R:Bertani, I.; Coglievina, M.; Zaccaria, P.; Klima, R.; Bruschi, C.V.
Yeast 11, 1187-1194, 1995
A:Title: The sequence of an 11.1 kb fragment on the left arm of Saccharomyces cerevisi
A:Reference number: S59235; MUID:96109931; PMID:8619317
A:Accession: S59238
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1528 <BEW>
A:Cross-references: UNIPARC:UPI000012DD97; EMBL:X84705; NID:g577853; PIDN:CAA59180.1;
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995
R:Bruschi, C.V.; Coglievina, M.; Bertani, I.; Klima, R.; Zaccaria, P.; Delneri, D.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64183
A:Accession: S64190
A:Molecule type: DNA

THIS PAGE BLANK (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:42:35 ; Search time 93.4006 Seconds
(without alignments)
1378.343 Million cell updates/sec

Title: US-10-635-977-2_COPY_73_365

Perfect score: 1560

Sequence: 1 EDITSADAVDLTEAWED.....LCAQVQEDTIKAVDRSCDI 293

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*
- 9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1560	100.0	293	8	ADJ93366 Human BGS
2	1560	100.0	541	8	ADJ93365 Human tub
3	1560	100.0	541	8	ADJ93358 Human BGS
4	1247	79.9	242	8	ADJ93360 Human BGS
5	908	58.2	352	3	AAB43005 Human ORF
6	908	58.2	352	8	AAM39450 Human pol
7	908	58.2	352	8	ADJ93457 Human HOT
8	908	58.2	352	8	ABM80420 Tumour-as
9	908	58.2	399	6	ABU11512 Human MDD
10	908	58.2	488	5	AAU74334 Human cyt
11	908	58.2	744	4	AAB94796 Human pro
12	905	58.0	326	7	ADM05524 Human pro
13	897	57.5	362	3	AAB58909 Breast an
14	817	52.4	292	8	ADJ93455 Human HOT
15	636	40.8	992	4	ABB65645 Drosophil
16	613	39.3	432	6	ABU92048 Human pro
17	515.5	33.0	719	4	ABB65541 Drosophil
18	480	30.8	362	4	AAM41236 Human pol
19	439	28.1	281	8	ADQ66614 Novel hum
20	374	24.0	566	4	ABQ05971 Novel hum
21	336	21.5	160	8	ADH45424 Human mol
22	330	21.2	347	6	ABU00150 Human nov
23	324	20.8	496	4	ABB64074 Drosophil
24	297.5	19.1	827	4	ABB60840 Drosophil

25	296	19.0	1281	4	AAM39105	Aam39105 Human pol
26	296	19.0	1304	4	AAM40891	Aam40891 Human pol
27	290.5	18.6	989	4	ABB60296	Abb60296 Drosophil
28	287	18.4	423	5	ABP43750	Abp43750 Ligase TT
29	287	18.4	423	8	ADJ93458	Adj93458 Human tub
30	281	18.0	673	8	ADQ66481	Adq66481 Novel hum
31	278	17.8	321	7	ADC33066	Adc33066 Human nov
32	278	17.8	393	7	ADC31411	Adc31411 Human nov
33	274	17.6	49	8	ADJ93373	Adj93373 Human BGS
34	272	17.4	487	4	ABB59096	Abb59096 Drosophil
35	271.5	17.4	519	4	AAM79305	Aam79305 Human pro
36	271.5	17.4	524	4	ABB22315	Abb22315 Protein #
37	271.5	17.4	592	7	ABB64837	Abb64837 Drosophil
38	270	17.3	917	4	ABB64837	Abb64837 Drosophil
39	256.5	16.4	439	7	ADB65504	Adb65504 Human pro
40	255.5	16.4	379	8	ADJ93456	Adj93456 Pig tubul
41	250.5	16.1	377	4	AAM79068	Aam79068 Human pro
42	250.5	16.1	377	5	ABB83472	Abb83472 Human cyt
43	250.5	16.1	377	7	ADE47756	Ade47756 Human NOV
44	250.5	16.1	377	8	ADJ79026	Adj79026 Human NOV
45	249.5	16.0	1138	8	ADJ71954	Adj71954 Human PMM

ALIGNMENTS

RESULT 1

ADJ93366	ID	ADJ93366	standard; protein; 293 AA.
XX	AC	ADJ93366;	
XX	DT	06-MAY-2004	(first entry)
XX	DE	Human BGS-42	protein-related TTL1 domain.
XX	KW	testis-specific tubulin tyrosine-ligase-like polypeptide;	
KW	KW	BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;	
KW	KW	neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;	
KW	KW	osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;	
KW	KW	anti-HIV; antibacterial; immunosuppressive; antiseborrheic;	
KW	KW	dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;	
KW	KW	tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;	
KW	KW	testicular disorder; testicular cancer; pulmonary disorder; lung cancer;	
KW	KW	gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;	
KW	KW	brain cancer; liver cancer; proliferative condition; testis; lung;	
KW	KW	small intestine; brain; lymph tissue; infertility; Cushing's syndrome;	
KW	KW	emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;	
KW	KW	Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;	
KW	KW	sepsis; acne; Sjogren's disease; scleroderma; human; TTL1 domain.	
XX	OS	Homo sapiens.	
XX	FN	WO2004005487-A2.	
XX	PD	15-JAN-2004.	
XX	PP	09-JUL-2003; 2003WO-US021605.	
XX	PR	09-JUL-2002; 2002US-0394725P.	
XX	PA	(BRIM) BRISTOL-MYERS SQUIBB CO.	
XX	PI	Feder JN, Wu S, Nelson TC;	
XX	XX	WPI; 2004-099381/10.	
XX	DR	New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,	
XX	PT	useful for preventing, treating or ameliorating a medical condition, e.g.	
XX	PT	aberrant cellular proliferation, reproductive disorders or testicular	
XX	XX	disorders.	
XX	PS	Disclosure; SEQ ID NO 14; 343pp; English.	

XX CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
 CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytostatic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
 CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 CC immunosuppressive, antiseborrheic or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant tubulin-
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of the TTL1
 CC domain of the human BGS-42 protein of the invention.
 XX CC
 SQ Sequence 293 AA;

Query Match 100.0%; Score 1560; DB 8; Length 293;
 Best Local Similarity 100.0%; Pred. No. 2.8e-160;
 Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EDIDTSADAVEDTEAEWEDLTQOYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 60
 DB 1 EDIDTSADAVEDTEAEWEDLTQOYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 60
 QY 61 IDGLRNIIWIKPAKSRGRDIVCMRVVEEILELAADHPLSRDNKNVQKYIETPLICD 120
 DB 61 IDGLRNIIWIKPAKSRGRDIVCMRVVEEILELAADHPLSRDNKNVQKYIETPLICD 120
 QY 121 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQFSLDKLDSAIHLCCNNAVQKYLKNDVGR 180
 DB 121 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQFSLDKLDSAIHLCCNNAVQKYLKNDVGR 180
 QY 181 SPLLPAHNMWTSRFBQYLQGRGAVGWSVYFPMKKATAHAMKVAQDHVPRKNSFEL 240
 DB 181 SPLLPAHNMWTSRFBQYLQGRGAVGWSVYFPMKKATAHAMKVAQDHVPRKNSFEL 240
 QY 241 YGADFVLGRDPRFWLIEINSPTHPSTPTVTAQCAQVQEDTIKVAVDKSCDI 293
 DB 241 YGADFVLGRDPRFWLIEINSPTHPSTPTVTAQCAQVQEDTIKVAVDKSCDI 293

RESULT 2
 ADJ93365
 ID ADJ93365 standard; protein; 541 AA.

AC ADJ93365;

XX 06-MAY-2004 (first entry)

XX Human tubulin tyrosine ligase protein consensus sequence SeqID13.

XX testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
 KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;

KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human.
 XX Homo sapiens.
 OS Synthetic.
 XX WO2004005487-A2.
 PN 15-JAN-2004.
 PD 09-JUL-2003; 2003WO-US021605.
 PF 09-JUL-2002; 2002US-0394725P.
 PR (BRIM) BRISTOL-MYERS SQUIBB CO.
 PA Feder JN, Wu S, Nelson TC;
 PI WPI; 2004-099381/10.
 DR N-PSDB; ADJ93364.
 XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
 PT useful for preventing, treating or ameliorating a medical condition, e.g.
 PT aberrant cellular proliferation, reproductive disorders or testicular
 XX disorders.
 XX Example 4; SEQ ID NO 13; 343pp; English.

PS This invention relates to a novel testis-specific tubulin tyrosine-ligase
 XX -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytostatic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
 CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 CC immunosuppressive, antiseborrheic or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant tubulin-
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of the
 CC tubulin tyrosine ligase protein consensus sequence which was used in the
 CC exemplification of the invention.

XX Sequence 541 AA;

Query Match 100.0%; Score 1560; DB 8; Length 541;
 Best Local Similarity 100.0%; Pred. No. 7e-160;
 Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDIDTSADAVEDTEAEWEDLTQOYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 60
 DB 73 EDIDTSADAVEDTEAEWEDLTQOYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 132
 QY 61 IDGLRNIIWIKPAKSRGRDIVCMRVVEEILELAADHPLSRDNKNVQKYIETPLICD 120
 DB 133 IDGLRNIIWIKPAKSRGRDIVCMRVVEEILELAADHPLSRDNKNVQKYIETPLICD 192

QY 121 TKFDIRQWFLVTDWNPNTIWFYKESYLRFSTQSFSLDKLDSAIHLNNNAVQKYLKNDVGR 180
 Db 193 TKFDIRQWFLVTDWNPNTIWFYKESYLRFSTQSFSLDKLDSAIHLNNNAVQKYLKNDVGR 252
 QY 181 SPLLPAHNWMTSTRFQEYLQGRGAVGWSVIYPSMKKATAHAKMVAQDHVPRKNSFEL 240
 Db 253 SPLLPAHNWMTSTRFQEYLQGRGAVGWSVIYPSMKKATAHAKMVAQDHVPRKNSFEL 312
 QY 241 YGADFVLGRDRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDRSCDI 293
 Db 313 YGADFVLGRDRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDRSCDI 365

RESULT 3

ADJ93358
 ID ADJ93358 standard; protein; 541 AA.

XX AC ADJ93358;

XX DT 06-MAY-2004 (first entry)

XX DE Human BGS-42 protein sequence SeqID2.

XX testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
 KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
 KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human.

OS Homo sapiens.

XX WO2004005487-A2.

XX 15-JAN-2004.

XX 09-JUL-2003; 2003WO-US021605.

XX 09-JUL-2002; 2002US-0394725P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Feder JN, Wu S, Nelson TC;

XX WPI: 2004-099381/10.

DR N-PSDB; ADJ93357.

XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
 PT useful for preventing, treating or ameliorating a medical condition, e.g.
 PT aberrant cellular proliferation, reproductive disorders or testicular
 PT disorders.

PS Claim 5; SEQ ID NO 2; 343pp; English.

XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
 CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytosolic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
 CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 CC immunosuppressive, antiseborrheic or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a

CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of the human
 CC BGS-42 protein of the invention.

XX SQ Sequence 541 AA;

Query Match 100.0%; Score 1560; DB 8; Length 541;
 Best Local Similarity 100.0%; Pred. No. 7e-160;
 Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDITSADAVEDLTEAEWEDLTQQYYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 60
 Db 73 EDITSADAVEDLTEAEWEDLTQQYYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 132
 QY 61 IDGLRNIIWIKPAKSRGRDIVCMRVEILEAADHPLSRDNKMWVQKYLETPLLICD 120
 Db 133 IDGLRNIIWIKPAKSRGRDIVCMRVEILEAADHPLSRDNKMWVQKYLETPLLICD 192
 QY 121 TKFDIRQWFLVTDWNPNTIWFYKESYLRFSTQSFSLDKLDSAIHLNNNAVQKYLKNDVGR 180
 Db 193 TKFDIRQWFLVTDWNPNTIWFYKESYLRFSTQSFSLDKLDSAIHLNNNAVQKYLKNDVGR 252
 QY 181 SPLLPAHNWMTSTRFQEYLQGRGAVGWSVIYPSMKKATAHAKMVAQDHVPRKNSFEL 240
 Db 253 SPLLPAHNWMTSTRFQEYLQGRGAVGWSVIYPSMKKATAHAKMVAQDHVPRKNSFEL 312
 QY 241 YGADFVLGRDRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDRSCDI 293
 Db 313 YGADFVLGRDRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDRSCDI 365

RESULT 4

ADJ93360
 ID ADJ93360 standard; protein; 242 AA.

XX AC ADJ93360;

XX DT 06-MAY-2004 (first entry)

XX DE Human BGS-42 protein sequence SeqID4.

XX testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
 KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
 KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human.

OS Homo sapiens.

XX WO2004005487-A2.

XX

PD 15-JAN-2004.
XX
XX
PF 09-JUL-2003; 2003WO-US021605.
XX
XX PR 09-JUL-2002; 2002US-0394725P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
PA
XX
XX
PI Feder JN, Wu S, Nelson TC;
XX
XX WPI; 2004-099381/10.
XX
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
XX Disclosure; SEQ ID NO 4; 343pp; English.
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiaesthetic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of the human
XX BGS-42 protein (partial sequence) of the invention.
XX
XX Sequence 242 AA;

Query Match 79.9%; Score 1247; DB 8; Length 242;
Best Local Similarity 100.0%; Pred. No. 2.2e-126; Indels 0; Gaps 0;
Matches 233; Conservative 0; Mismatches 0;

QY 61 IDGLRNITWIKPAKSRGRDIVCMRVBEIELEAAADHPLSRDNKVVQRYETPLLICD 120
DB 1 IDGLRNITWIKPAKSRGRDIVCMRVBEIELEAAADHPLSRDNKVVQRYETPLLICD 60
QY 121 TKFDIRQWFLVTDWNPITWIFKESYLRFSTQRFSLDKLSAIIHLNNVQKYLKNDVGR 180
DB 61 TKFDIRQWFLVTDWNPITWIFKESYLRFSTQRFSLDKLSAIIHLNNVQKYLKNDVGR 120
QY 181 SPLLPANHMTSTFQVLOQRGAVWGSVYIPSMKKATAHMKVAQDHVPEPKNSFEL 240
DB 121 SPLLPANHMTSTFQVLOQRGAVWGSVYIPSMKKATAHMKVAQDHVPEPKNSFEL 180
QY 241 YGADFVLGRDPRPWLIEINSPTHPTPTVTAQCAQVQEDTIKVAVDKRSDDI 293
DB 181 YGADFVLGRDPRPWLIEINSPTHPTPTVTAQCAQVQEDTIKVAVDKRSDDI 233

RESULT 5

AAB43005

ID AAB43005 standard; protein; 352 AA.

XX

AC AAB43005;
XX
XX DT 08-FEB-2001 (first entry)
XX
XX DE Human ORFX ORF2769 polypeptide sequence SEQ ID NO:5538.
XX
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
XX Homo sapiens.
OS
XX WO200058473-A2.
XX
XX PD 05-OCT-2000.
XX
XX PF 31-MAR-2000; 2000WO-US008621.
XX
XX PR 31-MAR-1999; 99US-0127607P.
XX PR 02-APR-1999; 99US-0127636P.
XX PR 05-APR-1999; 99US-0127728P.
XX PR 30-MAR-2000; 2000US-00540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M;
PI
XX WPI; 2000-602362/57.
XX N-PSDB; AAC77214.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 11; Page 4720-4721; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
XX Sequence 352 AA;

Query Match 58.2%; Score 908; DB 3; Length 352;
Best Local Similarity 56.7%; Pred. No. 2.6e-89;
Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;

XX 09-JUL-2002; 2002US-0394725P.
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX Feder JN, Wu S, Nelson TC;
XX WPI; 2004-099381/10.
XX
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
XX Disclosure; SEQ ID NO 7; 343pp; English.
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of the human
CC HOTTU3 protein which is related to the invention. Note: The present
CC sequence does not appear in the specification but was obtained from
CC Genbank.
XX
XX Sequence 352 AA;

Query Match 58.2%; Score 908; DB 8; Length 352;
Best Local Similarity 56.7%; Pred. No. 2.6e-89;
Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;

QY 2 DITSDADVEDLTEAEWEDLTQQYSLVHGDATFISNRNYSFQCQALLNRITSVNPQTDI 61
DB 2 DIDKDLEAPLYLTPEGWSLFQRYQVVGHEGAEHLHLDTVQRCEDILQQLQAVVQIDM 61
QY 62 DGLRNIWIIPAAKSRGRDIVCMRVEEILELAAADHPLSRDNKVVQVKYIETPLLCIDT 121
DB 62 EGDNRNIWVPGAKSRGIGIMCDHLEEMKLVNGNPVVMKDGKVVQVKYIERPLLIFFT 121
QY 122 KFDIRQWFLTDWNPPLTIWYKESYLFSTQSRSLDKLSAHLCHNNAVOKYLNVDGRS 181
DB 122 KFDLRQWFLTDWNPPLTWVFRYSYIRFSTQPSFLKNLSDNSVHLCHNNSIOKHLNSCHRH 181
QY 182 PLLPAHNMWTSRFOEYLQGRGAVGVMSVITYSPMKKAIAHAMKVQAQDHVEPRKNSPELY 241
DB 182 PLLPPDNNWSQRQAHLEQMGAPNASTIIVFGMKDAVHALQTSQDTVQCRKASPELY 241
QY 242 GADPVLGRDPRPMLIEINSSPTHPSTFVTAQLCAQVQEDETIKVAVDRSCD 292
DB 242 GADPVFCEGDFQWPLIEINASPTMAPSTAVTARLCAGVQADTLRVVIDRMLD 292

ABM80420
ID ABM80420 standard; protein; 352 AA.
XX
AC ABM80420;
XX
DT 18-NOV-2004 (first entry)
XX
DE Tumour-associated antigenic target (TAT) polypeptide PRO80956, SEQ:1058.
XX
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic.
XX
OS Homo sapiens.
XX
PN WO2004030615-A2.
XX
PD 15-APR-2004.
XX
PF 29-SEP-2003; 2003WO-US028547.
XX
PR 02-OCT-2002; 2002US-0414971P.
XX
PA (GETH) GENENTECH INC.
XX
PI Wu TD, Zhang Z, Zhou Y;
XX
XX WPI; 2004-347921/32.
XX N-PSDB; ACN37881.
XX
XX New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX
PS Claim 12; SEQ ID NO 1058; 7273pp; English.
XX
XX The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT polypeptide of the invention
XX
XX Sequence 352 AA;

Query Match 58.2%; Score 908; DB 8; Length 352;
Best Local Similarity 56.7%; Pred. No. 2.6e-89;
Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;

QY 2 DITSDADVEDLTEAEWEDLTQQYSLVHGDATFISNRNYSFQCQALLNRITSVNPQTDI 61
DB 2 DIDKDLEAPLYLTPEGWSLFQRYQVVGHEGAEHLHLDTVQRCEDILQQLQAVVQIDM 61
QY 62 DGLRNIWIIPAAKSRGRDIVCMRVEEILELAAADHPLSRDNKVVQVKYIETPLLCIDT 121
DB 62 EGDNRNIWVPGAKSRGIGIMCDHLEEMKLVNGNPVVMKDGKVVQVKYIERPLLIFFT 121
QY 122 KFDIRQWFLTDWNPPLTIWYKESYLFSTQSRSLDKLSAHLCHNNAVOKYLNVDGRS 181
DB 122 KFDLRQWFLTDWNPPLTWVFRYSYIRFSTQPSFLKNLSDNSVHLCHNNSIOKHLNSCHRH 181
QY 182 PLLPAHNMWTSRFOEYLQGRGAVGVMSVITYSPMKKAIAHAMKVQAQDHVEPRKNSPELY 241
DB 182 PLLPPDNNWSQRQAHLEQMGAPNASTIIVFGMKDAVHALQTSQDTVQCRKASPELY 241
QY 242 GADPVLGRDPRPMLIEINSSPTHPSTFVTAQLCAQVQEDETIKVAVDRSCD 292
DB 242 GADPVFCEGDFQWPLIEINASPTMAPSTAVTARLCAGVQADTLRVVIDRMLD 292

```

Db 62 EGDNRNIWIKPGAKSRGIMCMHLEMLKLVGNPNVVMKDGKVVQKYEIRPLLIQGT 121
QY 122 KFDIRQWFLVTDWNPITWIFYKESYLFSTQSFSLDKLDSAIHLGNNAVQKYLKNDVGRS 181
Db 122 KFDLRQWFLVTDWNPITWIFYRDSYIRFSTQSPFLKNDNSVHLCNNSIQKHLNSCHRH 181
QY 182 PLLPAHNNWTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFELY 241
Db 182 PLLPPDNMWSQRFOAHLQEMGAPNAWSTIIIVPGMKDAVIHALQTSQDVTQCRKASFELY 241
QY 242 GADFVLGRDRPRLWLIENSSPTMHPSTVTAQLCAOVQEDTIKVAVDKSCD 292
Db 242 GADFVGEDFPQWLIENASPTMAPSTAVTARLCAGVQADTLKRVVDRMLD 292

RESULT 9
ABU11512
ID ABU11512 standard; protein; 399 AA.
XX
AC ABU11512;
XX
DT 12-FEB-2003 (first entry)
DE Human MDDT polypeptide SEQ ID 459.
XX
KW MDDT; human; disease detection and treatment molecule polypeptide;
anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;
KW haemostatic; nephrotropic; antianaemic; antipsoriatic; hepatotropic;
KW gene therapy; protein replacement therapy; cell proliferative disorder;
KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; myeloma; sarcoma;
KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;
KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;
KW psoriasis; hepatitis.
XX
OS Homo sapiens.
XX
PN WO200279449-A2.
XX
PD 10-OCT-2002.
XX
PF 27-MAR-2002; 2002WO-US009944.
XX
PR 28-MAR-2001; 2001US-0279619P.
PR 29-MAR-2001; 2001US-0280067P.
PR 29-MAR-2001; 2001US-0280068P.
PR 16-MAY-2001; 2001US-0291280P.
PR 17-MAY-2001; 2001US-0291829P.
PR 17-MAY-2001; 2001US-0291849P.
PR 19-JUN-2001; 2001US-0299428P.
PR 20-JUN-2001; 2001US-0299776P.
PR 20-JUN-2001; 2001US-0300001P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;
PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI Flores V, Marwaha R, Lo A, Ian RY, Urashka ME;
XX
DR WPI; 2003-058431/05.
DR N-PSDB; ABX34502.
XX
PT New purified disease detection and treatment molecule proteins and
PT polynucleotides, useful for diagnosing, treating or preventing cancers
PT (e.g. leukaemia or sarcoma), anaemia, Crohn's disease, AIDS, osteoporosis
PT or hepatitis.
XX
PS Claim 27; SEQ ID NO 459; 339pp + Sequence Listing; English.
XX
XX This invention describes a novel disease detection and treatment molecule
CC polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,

```

```

CC osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,
CC antianaemic, antipsoriatic and hepatotropic activity. The polynucleotides
CC and the polypeptides of the invention can be used for gene therapy, of
CC protein replacement therapy and are useful for treating a variety of
CC diseases or conditions. These polypeptides or polynucleotides are
CC particularly useful for diagnosing, treating or preventing cell
CC proliferative disorders (e.g. cancers including adenocarcinoma,
CC leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's
CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's
CC syndromes, inflammation, osteoporosis, thrombocytopaenia, psoriasis or
CC hepatitis. ABU11450-ABU11845 represent the MDDT polynucleotides encoded
CC by ABU11450-ABU11845, described in the disclosure of the invention. NOTE:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 399 AA;

```

```

Query Match 58.2%; Score 908; DB 6; Length 399;
Best Local Similarity 56.7%; Pred. No. 3.1e-89;
Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;

QY 2 DDTSDADAVEDLTEAEWEDLTQOYYSLVHGDFAFISNRNYSFQCAALLNRITSVNPQTDI 61
Db 49 DIDKLEAPLYLTPEGWSLFLQRYQVYVHEGAELRLDPTQVQRCEDILQQLQAVVPQIDM 108

QY 62 DGLRNIWIKPAKSRGRDIVCMRDVREILEAAADHPLSRDNKVVQKYEIRPLLIQDT 121
Db 109 EGDNRNIWIKPGAKSRGIMCMHLEMLKLVGNPNVVMKDGKVVQKYEIRPLLIQDT 168

QY 122 KFDIRQWFLVTDWNPITWIFYKESYLFSTQSFSLDKLDSAIHLGNNAVQKYLKNDVGRS 181
Db 169 KFDLRQWFLVTDWNPITWIFYRDSYIRFSTQSPFLKNDNSVHLCNNSIQKHLNSCHRH 228

QY 182 PLLPAHNNWTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFELY 241
Db 229 PLLPPDNMWSQRFOAHLQEMGAPNAWSTIIIVPGMKDAVIHALQTSQDVTQCRKASFELY 288

QY 242 GADFVLGRDRPRLWLIENSSPTMHPSTVTAQLCAOVQEDTIKVAVDKSCD 292
Db 289 GADFVGEDFPQWLIENASPTMAPSTAVTARLCAGVQADTLKRVVDRMLD 339

```

```

RESULT 10
AAU74334
ID AAU74334 standard; protein; 488 AA.
XX
AC AAU74334;
XX
DT 12-MAR-2002 (first entry)
DE Human cytoskeleton-associated protein (CYSKP) #5.
XX
KW Human; cytoskeleton-associated protein; CYSKP; autoimmune disorder;
KW cell proliferative disorder; inflammatory disorder; prion disease;
KW vesicle trafficking disorder; gastrointestinal disorder; muscle disorder;
KW neurological disorder; cell motility disorder; reproductive disorder;
KW spinal cord disease; central nervous system disorder; mental disorder;
KW gene therapy; cancer.
XX
OS Homo sapiens.
XX
PN WO200185942-A2.
XX
PD 15-NOV-2001.
XX
PF 03-MAY-2001; 2001WO-US014355.
XX
PR 05-MAY-2000; 2000US-0201960P.
PR 08-MAY-2000; 2000US-0202729P.
PR 05-JUN-2000; 2000US-0209705P.
PR 07-JUN-2000; 2000US-0210149P.
PR 21-JUN-2000; 2000US-0213215P.

```


XX (INCY-) INCYTE GENOMICS INC.

XX Yue H, Tang YT, Au-Young J, Lu DAM, Baughn MR, Hallman JL;

XX Azimzai Y, Lal P, Yao MG, Bandman O, Burford N, Batra S, Kearney L;

XX Policky JL;

XX WPI; 2002-062248/08.

XX N-PSDB; AAS99894.

XX New cytoskeleton-associated proteins and polynucleotides, useful for

XX diagnosing, preventing and treating cell proliferative, autoimmune,

XX inflammatory, neurological, cell motility, reproductive and muscle

XX disorders.

XX Claim 1; Page 130-131; 194pp; English.

XX The invention relates to human cytoskeleton-associated polypeptides

XX (CYSKP) and their associated polynucleotide sequences. The sequences are

XX useful in the treatment of disorders associated with overexpression or

XX underexpression of CYSKP in a patient. The disorders include cell

XX proliferative disorders (such as cancer, actinic keratosis,

XX arterio-sclerosis, cirrhosis, hepatitis and psoriasis),

XX autoimmune/inflammatory disorders (such as, asthma, atherosclerosis,

XX osteoporosis, Crohn's disease, rheumatoid arthritis, diabetes mellitus

XX and anaemia), vesicle trafficking disorders (such as

XX hypercholesterolaemia, diabetes insipidus, Grave's disease and goitre),

XX gastrointestinal disorders, prion diseases, neurological disorders (such

XX as epilepsy, stroke, cerebral neoplasms, Alzheimer's disease,

XX Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis

XX and other motor neuron disorders), cell motility disorders, reproductive

XX disorders (such as endometriosis and polycystic ovary syndrome), muscle

XX disorders (such as myocarditis, migraine, hypertension, hypoglycaemia,

XX myocardial infarction, epilepsy and muscular dystrophy), spinal cord

XX diseases, central nervous system disorders (such as Down syndrome and

XX cerebral palsy) and mental disorders (such as anxiety and schizophrenia).

XX Sequences AAU74330-AAU74363 represent human CYSKP of the invention

XX

SQ Sequence 488 AA;

Query Match 58.2%; Score 908; DB 5; Length 488;

Best Local Similarity 56.7%; Pred. No. 4.2e-89;

Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;

QY 2 DITSADAVBDLTEAEWEDLTQOYSLVHGDAFINSRNYFSQCALLNRITSVNPQTDI 61

DB 2 DIDKDLAPLYLTPEGWSLFLQRYQVWHEGAEHLRLDTQVQRCEDILQQLQAVVQIDM 61

QY 62 DGLRNIWIKPAKSRGRDIVCMRDVEEILEAAADHPLSRDNKWWVQKYIETPLLCIDT 121

DB 62 EGDRIWIVKPGAKSRGIMCMCHLEEMLKLVNGNPVVMKDGKWWVQKYIERPLLIFFT 121

QY 122 KFDIROWFLVTDWNPLTIWFKESYLRFSQFSLKLDLSAHLNNAVOKYLNKDVGRS 181

DB 122 KFDLQWFLVTDWNPLTVWFYRDSYIRFSTQFSLKLDNSVHLNCSNIOKHLNSCHRH 181

QY 182 PLLPAHNMWTSRFOEYLQGRGAVMGSVIYFSMKKAIHAAMKVAQDHVPRKNSPELY 241

DB 182 PLLPDDNWSQSOFQALHGEWGPANWSTIIVGMDKDAVIAHALQTSQDTVQCRKASPELY 241

QY 242 GADFVLGRDPRPMIENSPTWHPSTPVTQAQLCAQVQEDTIKAVVDRSCD 292

DB 242 GADFVFGEDQPMLEINASPTWAPSTAVTARLCAGVQADTLRVYIDRMLD 292

RESULT 11

AAB94796

ID AAB94796 standard; protein; 744 AA.

XX AAB94796;

XX 26-JUN-2001 (first entry)

XX

DE Human protein sequence SEQ ID NO:15921.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1993; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-

XX length cDNAs defined in the specification, and for the detection and/or

XX diagnosis of the abnormality of the proteins encoded by the full-length

XX cDNAs.

XX Claim 8; SEQ ID NO 15921; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesising 5602 full-

XX length cDNAs defined in the specification. Where a primer set comprises:

XX (a) an oligo-dr primer and an oligonucleotide complementary to the

XX complementary strand of a polynucleotide which comprises one of the 5602

XX nucleotide sequences defined in the specification, where the

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

XX sequence and an oligonucleotide comprising a sequence complementary to a

XX polynucleotide which comprises a 3'-end sequence, where the

XX oligonucleotide comprises at least 15 nucleotides and the combination of

XX the 5'-end sequence/3'-end sequence is selected from those defined in the

XX specification. The primer sets can be used in antisense therapy and in

XX gene therapy. The primers are useful for synthesising polynucleotides,

XX particularly full-length cDNAs. The primers are also useful for the

XX detection and/or diagnosis of the abnormality of the proteins encoded by

XX the full-length cDNAs. The primers allow obtaining of the full-length

XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893

XX represent human amino acid sequences; and AAH13629 to AAH13632 represent

XX oligonucleotides, all of which are used in the exemplification of the

XX present invention

XX

SQ Sequence 744 AA;

Query Match 58.2%; Score 908; DB 4; Length 744;

Best Local Similarity 56.7%; Pred. No. 7.8e-89;

Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;

QY 2 DITSADAVBDLTEAEWEDLTQOYSLVHGDAFINSRNYFSQCALLNRITSVNPQTDI 61

DB 214 DIDKDLAPLYLTPEGWSLFLQRYQVWHEGAEHLRLDTQVQRCEDILQQLQAVVQIDM 273

QY 62 DGLRNIWIKPAKSRGRDIVCMRDVEEILEAAADHPLSRDNKWWVQKYIETPLLCIDT 121

DB 274 EGDRIWIVKPGAKSRGIMCMCHLEEMLKLVNGNPVVMKDGKWWVQKYIERPLLIFFT 333

QY 122 KFDIROWFLVTDWNPLTIWFKESYLRFSQFSLKLDLSAHLNNAVOKYLNKDVGRS 181

DB 334 KFDLQWFLVTDWNPLTVWFYRDSYIRFSTQFSLKLDNSVHLNCSNIOKHLNSCHRH 393

QY 182 PLLPAHNMWTSRFOEYLQROGRGAVWGSVIYPSMKKATAHAKVQAQDHVEPRKNSFELY 241
 DB 394 PLLPPDNWSSQRFQAHLEMGAPNAWSTIIIVPGMKDAVIHALQTSQDVTQCRKASFELY 453
 QY 242 GADFLVGRDRFRPWLIEINSSPTMHPSTPVTQAOLCAQVQEDTIKVAVDKSCD 292
 DB 454 GADFLVGEDFQPLWLEINASPMTAPSTAVTARLCAGVQADTLRVVIDRMLD 504

RESULT 12
 ADM05524
 ID ADM05524 standard; protein; 326 AA.
 XX
 AC ADM05524;
 XX
 DT 20-MAY-2004 (first entry)
 XX
 DE Human protein of the invention SEQ ID NO:4209.
 XX
 KW human; gene therapy; diagnostic marker; pharmaceutical.
 XX
 OS Homo sapiens.
 XX
 PN EP1347046-A1.
 XX
 PD 24-SEP-2003.
 XX
 PF 12-APR-2002; 2002EP-00008400.
 XX
 PR 22-MAR-2002; 2002JP-00137785.
 XX
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
 XX
 DR WPI; 2003-723558/69.
 DR N-PSDB; ADM03081.
 XX
 PT New polynucleotides and polypeptides are useful in gene therapy, for
 PT developing a diagnostic marker or medicines for regulating their
 PT expression and activity, or as a target of gene therapy.
 XX
 PS Claim 1; SEQ ID NO 4209; 305pp; English.
 XX
 CC The invention relates to a novel human polynucleotide and the encoded
 CC polypeptide. A polynucleotide of the invention may have a use in gene
 CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
 CC as a primer for synthesizing the polynucleotide or as a probe for
 CC detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
 CC useful in gene therapy, for developing a diagnostic marker or medicines
 CC for regulating their expression and activity, or as a target of gene
 CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
 CC are useful as pharmaceutical agents. The present sequence represents a
 CC protein sequence of the invention.
 XX
 SQ Sequence 326 AA;
 Query Match 58.0%; Score 905; DB 7; Length 326;
 Best Local Similarity 56.7%; Pred. No. 4.9e-89;
 Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;
 QY 2 DIDTSADAVEDLTAEWEDLTQOYSLVHGDAFISNRNYSQCALNRTISVNPQTDI 61
 DB 2 DIDKDLAPLYLTPEGNSLFLORYQVHGEALHRLDTQVQRCEDIQLQQLQAVVPQDM 61
 QY 62 DGLRNWIKPAKSGRGRDVCMDRVEEILEAAADHPLSRDNKWKVQKIETPLLCDT 121
 DB 62 EGDRIWIVKPGAKSGRGIMCMDHLEMLKLVNGNVPVMKDGKWKVQKIETPLIFGT 121
 QY 122 KFDIRQWFLVTDWNPPLTWPFYKESVLRSTQRFSLKLNLSVHLCCNSVHLKNSCHRH 181

DB 122 KFDLRQWFLVTDWNPPLTWPFYRDSYRSTQRFSLKLNLSVHLCCNSVHLKNSCHRH 181
 QY 182 PLLPAHNMWTSRFOEYLQROGRGAVWGSVIYPSMKKATAHAKVQAQDHVEPRKNSFELY 241
 DB 182 PLLPPDNWSSQRFQAHLEMGAPNAWSTIIIVPGMKDAVIHALQTSQDVTQCRKASFELY 241
 QY 242 GADFLVGRDRFRPWLIEINSSPTMHPSTPVTQAOLCAQVQEDTIKVAVDKSCD 292
 DB 242 GADFLVGEDFQPLWLEINASPMTAPSTAVTARLCAGVQADTLRVVIDRMLD 292

RESULT 13
 AAB58909
 ID AAB58909 standard; protein; 362 AA.
 XX
 AC AAB58909;
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 617.
 XX
 KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KW neoplastic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antiinflammatory; antiulcer; vulnery; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200055173-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US005881.
 XX
 PR 12-MAR-1999; 99US-0124270P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-611515/58.
 DR N-PSDB; AAF21812.
 XX
 PT New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention, treatment
 PT and diagnosis of cancer, immune disorders, cardiovascular disorders and
 PT neurological diseases.
 XX
 PS Claim 11; Page 1056-1057; 1299pp; English.
 XX
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive; neoplastic;
 CC neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
 CC antiinflammatory; antiulcer; vulnery; anticonvulsant; antibacterial;
 CC antifungal; antiparasitic and cardiant activity. The polynucleotide and
 CC protein sequences are used in the diagnosis of cancer, particularly
 CC breast and ovarian cancer. The nucleic acid sequences, proteins, agonists
 CC and agonists may also be used in the diagnosis, prevention and treatment
 CC of immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC cardiovascular disorders such as myocardial ischaemia; wound healing;
 CC neurological diseases such as cerebral anoxia and epilepsy; and
 CC infectious diseases

```
XX SQ Sequence 362 AA;
Query Match 57.5%; Score 897; DB 3; Length 362;
Best Local Similarity 57.5%; Pred. No. 4.2e-88;
Matches 161; Conservative 47; Mismatches 72; Indels 0; Gaps 0;
QY 13 LTAEWEDLTQYYSLVHGDAFTNSGRNYSFQOALLNRITSVNPQTDIDGLNIWIKP 72
DB 30 LTPEGHSLFLQYQYVHGAELRHLDTVQRCEDILQQLQAVVPQIDMEGDRNIWIKP 89
QY 73 AAKSRGRDIVCMRDRVEIELEAAADHPLSRDNKWWVQKYEITETPLLLICDTKFDIROWFLVT 132
DB 90 GANSRGRGIMCMHLEMLKLVNGVPMVMDGKGVVQKYEIRPELLFGTKFDIROWFLVT 149
QY 133 DMNPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGRSPLLPANHMTS 192
DB 150 DMNPLTIWFYRDSYIRFSTQRFSLDKLDSVHLCCNNSIQHLENSCHRHPELLPPDNWSS 209
QY 193 TRFOEYLQROGRGAVGWSVITYPSMKKAIAHAMKVAQDHVEPRKNSFELGADFLGRDFR 252
DB 210 QRFQALHLEMGAPNAWSTIIVPGMKDAVIHALQTSQDTVQCRKASFELYGADVFGEQDFQ 269
QY 253 PWLIENSSPTMHPSTPTVTAOLCAQVQEDTIKVAVDKSCD 292
DB 270 PWLIENASPTMAPSTAVTARLCAGVQADTLRVVDIRXLD 309

RESULT 14
ADJ93455
ID ADJ93455 standard; protein; 292 AA.
XX AC ADJ93455;
XX XX
XX 06-MAY-2004 (first entry)
XX DE Human HOTTL protein sequence SeqID2.
XX testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-gen; gastrointestinal-gen;
KW neuroprotective; endocrine-gen; antiinflammatory-gen; anabolic; hypertensive;
KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; HOTTL.
XX OS Homo sapiens.
XX WO2004005487-A2.
XX 15-JAN-2004.
XX 09-JUL-2003; 2003WO-US021605.
XX 09-JUL-2002; 2002US-0394725P.
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX Feder JN, Wu S, Nelson TC;
XX WPI; 2004-099381/10.
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
XX useful for preventing, treating or ameliorating a medical condition, e.g.
XX aberrant cellular proliferation, reproductive disorders or testicular
XX disorders.
```

```
XX PS Disclosure; SEQ ID NO 5; 343pp; English.
XX XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of the human
CC HOTTL protein which is related to the invention. Note: The present
CC sequence does not appear in the specification but was obtained from
CC Genbank.
XX SQ Sequence 292 AA;
Query Match 52.4%; Score 817; DB 8; Length 292;
Best Local Similarity 62.5%; Pred. No. 1.5e-79;
Matches 145; Conservative 37; Mismatches 50; Indels 0; Gaps 0;
QY 61 IDGRNIWIKPAKSRGRDIVCMRDRVEIELEAAADHPLSRDNKWWVQKYEITPLLICD 120
DB 1 MEGRDFWIVKPGAKSRGRGIMCMHLEMLKLVNGVPMVMDGKGVVQKYEIRPLIFG 60
QY 121 TKEDIRQWFLVTDWNPITTWYKESYLRFSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 180
DB 61 TKEDLRQWFLVTDWNPITTWYKESYLRFSTQRFSLDKLDSVHLCCNNSIQHLENSCHR 120
QY 181 SPLLPAHNNMTSTRFOEYLQROGRGAVGWSVITYPSMKKAIAHAMKVAQDHVEPRKNSFEL 240
DB 121 HPLLPPDNWSSQRFQALHLEMGAPNAWSTIIVPGMKDAVIHALQTSQDTVQCRKASFEL 180
QY 241 YGADFVLGRDFRPLWLEINSPTMHPSTPTVTAOLCAQVQEDTIKVAVDKSCD 292
DB 181 YGADFVFGEDFQPLWLEINASPTMAPSTAVTARLCAGVQADTLRVVDIRXLD 232

RESULT 15
ABB65645
ID ABB65645 standard; protein; 992 AA.
XX AC ABB65645;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 23727.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
PD
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:59:01 ; Search time 12.0669 Seconds
(without alignments)
1364.812 Million cell updates/sec

Title: US-10-635-977-2

Perfect score: 2854

Sequence: 1 MASSILKWVSHQCSRSSR.....LRGLKTAEGALRPPPGKGKS 541

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 180808 seqs, 30441898 residues

Total number of hits satisfying chosen parameters: 180808

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New: *

- | | | | | |
|----|-----------------------------|-----|-----|------|
| 1: | SIDS5\ptodata\2\pubpaa\US08 | NEW | PUB | dep. |
| 2: | SIDS5\ptodata\2\pubpaa\US06 | NEW | PUB | dep. |
| 3: | SIDS5\ptodata\2\pubpaa\US07 | NEW | PUB | dep. |
| 4: | SIDS5\ptodata\2\pubpaa\PCT | NEW | PUB | dep. |
| 5: | SIDS5\ptodata\2\pubpaa\US09 | NEW | PUB | dep. |
| 6: | SIDS5\ptodata\2\pubpaa\US10 | NEW | PUB | dep. |
| 7: | SIDS5\ptodata\2\pubpaa\US11 | NEW | PUB | dep. |
| 8: | SIDS5\ptodata\2\pubpaa\US60 | NEW | PUB | dep. |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	280.5	9.8	439	7	US-11-072-512-3658	Sequence 3658, App
2	271.5	9.5	592	7	US-11-072-512-3371	Sequence 3371, App
3	183	6.4	497	7	US-11-072-512-3621	Sequence 3621, App
4	127.5	4.5	704	6	US-10-330-773-663	Sequence 663, App
5	108	3.8	713	6	US-10-330-773-668	Sequence 668, App
6	107.5	3.8	402	6	US-10-649-591-18	Sequence 18, App
7	103	3.6	710	7	US-11-072-512-3402	Sequence 3402, App
8	101.5	3.6	723	7	US-11-072-512-2246	Sequence 2246, App
9	100	3.5	354	7	US-11-072-512-3151	Sequence 3151, App
10	100	3.5	756	7	US-11-072-512-2505	Sequence 2505, App
11	99.5	3.5	423	7	US-11-000-463-375	Sequence 375, App
12	99.5	3.5	1892	7	US-11-075-185-6	Sequence 6, Appli
13	98.5	3.5	386	7	US-11-096-568A-21828	Sequence 21828, A
14	98.5	3.5	2760	7	US-11-124-367A-444	Sequence 444, App
15	98.5	3.5	2803	7	US-11-124-367A-442	Sequence 442, App
16	98.5	3.5	2803	7	US-11-124-367A-445	Sequence 445, App
17	98.5	3.5	2984	7	US-11-124-367A-443	Sequence 443, App
18	98.5	3.5	3027	7	US-11-124-367A-441	Sequence 441, App
19	98	3.4	395	6	US-10-330-773-661	Sequence 661, App
20	97.5	3.4	1137	7	US-11-087-099-5799	Sequence 5799, App
21	96	3.4	193	7	US-11-096-568A-24769	Sequence 24769, A
22	95.5	3.3	1560	7	US-11-059-982-1	Sequence 1, Appli
23	95	3.3	1767	6	US-10-995-561-911	Sequence 911, App
24	95	3.3	1767	6	US-10-995-561-914	Sequence 914, App
25	95	3.3	2472	7	US-11-203-806A-2	Sequence 2, Appli

```
Db      125  EREAGLEAAKCDFFPKTTEPCPEYHLFVEERKKNFGITWIMKPVARSOGKGLFPRUK 184
Qy      161  ETELEAAADHPLSRDNK-----WVQKYIETPLLICDTKFDIRQWFLVTDWNPWTIW 213
Db      185  DIVDW-RKOTRSDDQKDDIPVENYQAQYIENPYLIGGRKFDLRYVVLMS-----V 236
Qy      214  YKESYLRFTQRFSLDKLSAHLNNAVQKYLKNDVGRSPLLPAIN-----MWTSTRFQ 268
Db      237  FAECLLWSGHR-----QDVHLTNVAVQK-----TSPDYHPKKGCKWTLQRF 279
Qy      269  EYL-QROGRGAVGWSVIYPSMKKAIHAAMKVAQDHVEPRKNSFELYGADFVLGRDPRWL 327
Db      280  QYLASHKHGPEAV--ETLFRDIDNIFVKSQSVQKVIISDKHCFELYGYDILIDQDLKPL 337
Qy      328  IEINSSPTMHPSPFTVTAQLCAQVOEDTIKVA-----VDRSCDIGNFELLWRQPVE--- 378
Db      338  LEVNASPSLTASSQEDYELKTCLEDTLHVDMEARLTGREKRVGGFDLMWNDGPFVSREE 397
Qy      379  -PPFSGSLCVAGVSV-----RRARRQLVPCNL-----KASA 411
Db      398  GAPDLGSMGNFVNTNTHLGCVNDKQKQLRQLF--CSLQVQKQASS 439
```

RESULT 2

```
US-11-072-512-3371
; Sequence 3371, Application US/11072512
; Publication No. US20060029945A1
```

GENERAL INFORMATION:

```
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: TRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cdna
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3371
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
US-11-072-512-3371
```

```
Query Match      9.5%; Score 271.5; DB 7; Length 592;
Best Local Similarity 29.8%; Pred. No. 4.1e-15;
Matches 72; Conservative 43; Mismatches 80; Indels 47; Gaps 8;

Qy      113  YFSQCQALLNRITSVNPTQDIDGLRNITWIKPAKSRGRDIVCMRVEEILELAADHPL 172
Db      195  YFQERQMLGTG-----HSYWICKPAELSRGRGILIFSDFKDFI----- 232
Qy      173  SRDNKVVQKYIETPLLICDTKFDIRQWFLVTDWNPWTIWFKESYLRFSFORSLDKL- 231
```

```
Db      233  -FDDMYIVQKYINPILLIGRYKCDLRIYVCVTGFKPLTIYVYQEBGLVRPATEKFDLSNLQ 291
Qy      232  DSAIHLNNAVQK-----YLNKDVGRSPLLAHNMTSTRQEYLRQGRGAVW---GS 282
Db      292  NNVAHLTNSSINKSGASYEKIKEVIGHG-----CKWTLRSRFSYLR-----WDVDDL 339
Qy      283  VIYPSMKKAIHAAMKVAQDHVEPRKNSFELYGADFVLGRDPRFWLIEINSSP--TMHPST 340
Db      340  LLWKIHRMVIITILAIAPSVFPAANCFELFGDILIDNLPWLLLEVNSPALIDCST 399
Qy      341  PV 342
Db      400  DV 401
```

RESULT 3

```
US-11-072-512-3621
; Sequence 3621, Application US/11072512
; Publication No. US20060029945A1
```

GENERAL INFORMATION:

```
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: TRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cdna
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3621
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
US-11-072-512-3621
```

```
Query Match      6.4%; Score 183; DB 7; Length 404;
Best Local Similarity 25.5%; Pred. No. 8.6e-08;
Matches 82; Conservative 38; Mismatches 118; Indels 84; Gaps 17;

Qy      6   LKVVVSHQSCRSRSRKPDRQREAGSSDLSSRQDAENAEAK-----LRGLPGQLVDIA 59
Db      119  LKW-----CEVKSRSYSGSFRE--GRQLLYQLPNNKLLTTKIGLLSTLRG--RARAMSRA 169
Qy      60  CKVQAYLQGLEHEDITSADAVEDLTEAWEDLTQYYSLVHGDAFISNRNYSFSCQA 119
Db      170  SKVPGGVQARLEK---DAAAPALEDL---PW-----TSPGYLRPQRV 205
Qy      120  LLNRITSVNPTQ---DIDGLR-----NIWIKPAKSRGRDIVCMRVEEILELA 166
Db      206  L--RMEEFFETVRLDLKHEREAFFTLFDETOIWKPTASNQCKGIFLLNQEEVAALQ 263
Qy      167  A-----ADHPLSRDNKW-----VVKYIETPLLICDTKFDIRQWFLVTDWNPWTIWFK 215
Db      264  AKTRSMEDDPIHHKTPFRGQARVWQRYIQNPLLDVGRKFDVRSYLLIACCTPYMI-PFG 322
```



```
; GENERAL INFORMATION:
; APPLICANT: Tang, Y Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Chen, Rui-hong
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Zhou, Ping
; APPLICANT: Cao, Yi-Cheng
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 785CIP4CN
; CURRENT APPLICATION NUMBER: US/11/000,463
; CURRENT FILING DATE: 2004-11-29
; PRIOR APPLICATION NUMBER: 10/291,265
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 09/922,279
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 375
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-000-463-375
```

```
Query Match 3.5%; Score 99.5; DB 7; Length 423;
Best Local Similarity 24.4%; Pred. No. 1.3;
Matches 89; Conservative 34; Mismatches 125; Indels 117; Gaps 22;

Qy 181 QKIETPLLCDFKFDIRQWFLVTD-----WNPLTIWYF---KSYLRFSTQSFSLDKLDS 233
Db 128 QKLFEKILKRDVAHKV---FATTSIKSFQQLNLYGFRKRROCTFRTFTRIFSAGELVS 184

Qy 234 AIHLNNNAVQKLYKNDVGRSPLLPAH-----NMWTSRFOYLQGRGAVWGSGVIYPSM 288
Db 185 ILNKLEFYCHPYFORD---SPHLLVRMKRRVGVKSAPRHQEDKPEAAGS-----231

Qy 289 KKAIAHAMKVAQDHVEPRKNSFELYGADFVLGRDRP-----WLIEINSSPTMHPSTPVT 344
Db 232 --CLAPADTEQODHTSPNEN-----DQVTPQHREPAQNPQTQIRSGSAP---PATPVW 279

Qy 345 QLCQAOVEDTIKAVDRSCDIGNPELLWRQVVEPPPPPSGSDLCVAGSVRRARRQVLVP 404
Db 280 PDSA-----VASDNS-----PVTQP-----AGEWSESGQAHTVP 309

Qy 405 CNLKASASLLDAQPKARGPSAMP-----DPAQ-----GP--PSPALQDRLGLKEEGLP 452
Db 310 A-----AVP-----GPNALPFLYVPGSPTQMSYGPVVVALPTASRSTLAMDTTGLP 355

Qy 453 LALLAP-----LRGAASGGAAQPTRTKAAGKV---ELPA-----CPCRHVDSQ--APNTGV 499
Db 356 APGMLPFCHLWVPVTLVAAGAAQP-----AASVMVFPPLPALHHCHPHSHRTSQYWPASDG 411

Qy 500 PVAQP 504
Db 412 PQAYP 416
```

RESULT 12
US-11-075-185-6

```
; Sequence 6, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 6
; LENGTH: 1892
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-11-075-185-6
```

```
Query Match 3.5%; Score 99.5; DB 7; Length 1892;
Best Local Similarity 20.4%; Pred. No. 11;
Matches 133; Conservative 70; Mismatches 221; Indels 229; Gaps 30;

Qy 20 RSKPRQREBA-----GSSDLSRQDAENAEAKRLGLP-----52
Db 1154 QSLARLSREBAAGDPVLVPFAWTVDVALYATGATELRARIALEQASG--GAPAVASLLL 1210

Qy 53 ---GOLVDIACKVCOAYLGQLEHEDITSADAVEDLTAEWEDL-----TQYYSLV 101
Db 1211 ADAHRTVATTGRVGSAAQTR-----SAASRAEPMYRVAVTVDVALEATWAEHVL 1265

Qy 102 HGDAFISNSRNYFSQCALNRIITSVNPQTDIDGLRNIMIKPAKSRGRDVCMDRVEE 161
Db 1266 GGDGALA-----AALGVRAAAGLP-LEALAD-----GAAAPR-----RL 1300

Qy 162 ILELAAD-----HPLSRDNKVVQKVIETPLLICDTKFIROWFLVTD-----W 206
Db 1301 VDLTAGDAGAVVAHVAAVVGALALVQGWLAAPQLAATELLVVTTRCAVATGDEGVDAL 1360

Qy 207 NPLTIWFKESYLRFSTQRFSLDKLSAHLCCNNAVQKLYKNDVGRSPLLPAHMMWTSTR 266
Db 1361 GPAVW-----GLLRATRAEYP-----DRAVRVL-----DVGREPLDGA-----1394

Qy 267 FQYLQROGRGAVWGSGVIYPSMKKAIAHAMKVAQ-----DHVEPRKNSFELYGA 315
Db 1395 ----LLRRALAA--GTEPELSVRSGEARAPRLREVRSSEPAAPATRLDPDGTAL-ITGG 1447

Qy 316 DFVLGRDRFRPWLIEINSSPTMH-----PSTPVTQALCAQVQ---EDTIKVAVD 360
Db 1448 TGBLGRHVAKHLV-----TAHGVRHLVLTSSRRGMDAPDAALVDELRAAGAATVDVA-- 1499

Qy 361 RSCDIGNFELLWQPVVEPPPPSG-----SDLCVAGSVVRRARRQVLVP-----403
Db 1500 -ACDAADAAAL--AAVVEALPAARPLTAVVHTTAGVLDSDSVTKLSAQRLARVLPKVDGA 1556

Qy 404 -----VCNLKASASLLD--AQPLKARGPSAMPDPAQ 432
Db 1557 FHLHELTKHAPLAAPVLFSSAAGTGLSGPGQANVAAANTFLDALASHLRARG-----1607

Qy 433 GPSPALQDRLGLKEEGLPLALLAPLRGAESGGAQPTRTKAAGKVELP-ACPCRHDV 491
Db 1608 ---VPMSLAWEFGWAQTGL-----GMTAHLGAADIARMKRKHGVVSMFVAGQLRLD 1655

Qy 492 ---SQAPNTGVPVPAQPAKSWDPNQLNAHPLFEPVLRGLKTAEGALRPPPGGKGS 541
Db 1656 RALAQAETLVPLALDLSLQKAGSNAGPVPPLRLGLVRAPAGRRTAASAAGA 1708
```

RESULT 13
US-11-096-568A-21828

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:55:16 ; Search time 77.9322 Seconds
(without alignments)
2900.544 Million cell updates/sec

Title: US-10-635-977-2
Perfect score: 2854
Sequence: 1 MASSILKVVVSHQSCSRSSR.....LRGLKTAEALRPPPGKXGS 541

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_Main:*
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2854	100.0	541	4	US-10-615-659-2
2	2854	100.0	541	4	US-10-635-977-2
3	2843	99.6	541	4	US-10-615-659-13
4	2843	99.6	541	4	US-10-635-977-13
5	1560	54.7	293	4	US-10-615-659-14
6	1560	54.7	293	4	US-10-635-977-14
7	1299	45.5	242	4	US-10-615-659-4
8	1299	45.5	242	4	US-10-635-977-4
9	1013	35.5	488	4	US-10-275-595A-5
10	969.5	34.0	362	3	US-09-925-298-617
11	969.5	34.0	362	4	US-10-102-806-617
12	956.5	33.5	352	4	US-10-615-659-7
13	956.5	33.5	352	4	US-10-635-977-7
14	956.5	33.5	352	5	US-10-756-149-5267
15	944.5	33.1	326	4	US-10-108-260A-4209
16	865.5	30.3	292	4	US-10-615-659-5
17	865.5	30.3	292	4	US-10-635-977-5
18	722.5	25.3	992	6	US-11-097-143-23727
19	536.5	18.8	719	6	US-11-097-143-23415
20	375.5	13.2	566	5	US-10-450-763-36330
21	328	11.5	496	6	US-11-097-143-13014
22	322	11.3	330	4	US-10-615-659-26
23	322	11.3	330	4	US-10-635-977-26
24	303.5	10.6	1226	5	US-10-756-149-5184
25	302	10.6	268	4	US-10-424-599-205823
26	299.5	10.5	423	4	US-10-615-659-8
27	299.5	10.5	423	4	US-10-635-977-8

28	297.5	10.4	827	6	US-11-097-143-9312	Sequence 9312, Ap
29	291	10.2	989	6	US-11-097-143-7680	Sequence 7680, Ap
30	282.5	9.9	487	6	US-11-097-143-4080	Sequence 4080, Ap
31	280.5	9.8	439	4	US-10-104-047-3658	Sequence 3658, Ap
32	274	9.6	49	4	US-10-615-659-21	Sequence 21, Appl
33	274	9.6	49	4	US-10-615-659-22	Sequence 22, Appl
34	274	9.6	49	4	US-10-635-977-21	Sequence 21, Appl
35	274	9.6	49	4	US-10-635-977-22	Sequence 22, Appl
36	271.5	9.5	524	3	US-09-864-761-38213	Sequence 38213, A
37	271.5	9.5	592	4	US-10-104-047-3371	Sequence 3371, Ap
38	270	9.5	917	6	US-11-097-143-21303	Sequence 21303, A
39	259	9.1	379	4	US-10-615-659-6	Sequence 6, Appl
40	259	9.1	379	4	US-10-635-977-6	Sequence 6, Appl
41	258	9.0	92	4	US-10-424-599-262294	Sequence 262294,
42	254.5	8.9	377	4	US-10-210-130-1118	Sequence 118, App
43	254.5	8.9	377	4	US-10-250-613-1	Sequence 1, Appl
44	237	8.3	42	4	US-10-615-659-24	Sequence 24, Appl
45	237	8.3	42	4	US-10-635-977-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1

US-10-615-659-2
; Sequence 2, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-2

Query Match	100.0%	Score 2854;	DB 4;	Length 541;
Best Local Similarity	100.0%	Pred. No. 3.1e-229;		
Matches 541;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MASSILKVVVSHQSCSRSSRSPKPRDQREAGSSDLSSRQDAENAEAKLRGLPGQLVDIAC	60	
Db	1	MASSILKVVVSHQSCSRSSRSPKPRDQREAGSSDLSSRQDAENAEAKLRGLPGQLVDIAC	60	
Qy	61	KVCQAYLGQLEHEDITSDADAVEDLTEAEWEDLTQYYSLVHGDAFISNSRNYFSCQAL	120	
Db	61	KVCQAYLGQLEHEDITSDADAVEDLTEAEWEDLTQYYSLVHGDAFISNSRNYFSCQAL	120	
Qy	121	LNRTSVNPNQDIDGLRNIIWKPAKSRGRDIVCMCDRVEELEAAADHPLSRNKKVV	180	
Db	121	LNRTSVNPNQDIDGLRNIIWKPAKSRGRDIVCMCDRVEELEAAADHPLSRNKKVV	180	
Qy	181	QKXIETPLLCIDTKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCNN	240	
Db	181	QKXIETPLLCIDTKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCNN	240	
Qy	241	AVQKYLKNDVGRSPLLPANNMTSTRFQYLRQGRGAVGWSVIYPSMKKAIAHAMKVAQ	300	
Db	241	AVQKYLKNDVGRSPLLPANNMTSTRFQYLRQGRGAVGWSVIYPSMKKAIAHAMKVAQ	300	
Qy	301	DHVEPRKNSFELYGADFVLGRDFRPWLIEINSSPTMHSTPTVTAQLCAQVQEDTIKVAVD	360	
Db	301	DHVEPRKNSFELYGADFVLGRDFRPWLIEINSSPTMHSTPTVTAQLCAQVQEDTIKVAVD	360	
Qy	361	RSCDIGNPELLWRQPVVPPPPFSGSDLCVAGSVRRARRQVLPVCNKLKASALLDAQPLK	420	

Db 361 RSCDIGNFELLWQPVVPPFSGSDLCVAGSVRRARROVLPVCNLKASASLLDAQPLK 420
Qy 421 ARGPSAMPDPAQPPSPALQDRLGLKEEKGPLALLAPLRGAESGAAQPTRTKAAGKV 480
Db 421 ARGPSAMPDPAQPPSPALQDRLGLKEEKGPLALLAPLRGAESGAAQPTRTKAAGKV 480
Qy 481 ELPACPCRHVDSQAPNTGVPVAQPAKSWDPNQLNAHPLEPVLRLGLKTAEGALRPPPGKG 540
Db 481 ELPACPCRHVDSQAPNTGVPVAQPAKSWDPNQLNAHPLEPVLRLGLKTAEGALRPPPGKG 540
Qy 541 S 541
Db 541 S 541

RESULT 2
US-10-635-977-2
; Sequence 2, Application US/10635977
; Publication No. US2004017113A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-2

Query Match 100.0%; Score 2854; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 3.1e-229;
Matches 541; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MASSILKWVSHQSCSRSSKPRDQREAEAGSSDLSSRQDAENAEAKLGLPGQLVDIAC 60
Db 1 MASSILKWVSHQSCSRSSKPRDQREAEAGSSDLSSRQDAENAEAKLGLPGQLVDIAC 60
Qy 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQQYYSLVHGDAFISNSRNYFSQCOAL 120
Db 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQQYYSLVHGDAFISNSRNYFSQCOAL 120
Qy 121 LNRTSVNPTDIDGLRNIWIIPAAKSRGRDIVCMRDVEEILELAADHPLSRDNKVV 180
Db 121 LNRTSVNPTDIDGLRNIWIIPAAKSRGRDIVCMRDVEEILELAADHPLSRDNKVV 180
Qy 181 QKYIETPLLCIDTKFDIRQWFLVTDWNPFTIWFYKESYLPSTQRFSLDKLDSAIHLCNN 240
Db 181 QKYIETPLLCIDTKFDIRQWFLVTDWNPFTIWFYKESYLPSTQRFSLDKLDSAIHLCNN 240
Qy 241 AVQKYLKNDVGRSPLLAHNNMTSTRFQEYLQROGRGAVGWSVIYPSMKKAIHAAMKVAQ 300
Db 241 AVQKYLKNDVGRSPLLAHNNMTSTRFQEYLQROGRGAVGWSVIYPSMKKAIHAAMKVAQ 300
Qy 301 DHVEPRKNSFELYGADFVLGRDFRPWLIEINSSPTMHPSTPTVAQLCAQVOEDTIKVAVD 360
Db 301 DHVEPRKNSFELYGADFVLGRDFRPWLIEINSSPTMHPSTPTVAQLCAQVOEDTIKVAVD 360
Qy 361 RSCDIGNFELLWQPVVPPFSGSDLCVAGSVRRARROVLPVCNLKASASLLDAQPLK 420
Db 361 RSCDIGNFELLWQPVVPPFSGSDLCVAGSVRRARROVLPVCNLKASASLLDAQPLK 420
Qy 421 ARGPSAMPDPAQPPSPALQDRLGLKEEKGPLALLAPLRGAESGAAQPTRTKAAGKV 480

Db 421 ARGPSAMPDPAQPPSPALQDRLGLKEEKGPLALLAPLRGAESGAAQPTRTKAAGKV 480
Qy 481 ELPACPCRHVDSQAPNTGVPVAQPAKSWDPNQLNAHPLEPVLRLGLKTAEGALRPPPGKG 540
Db 481 ELPACPCRHVDSQAPNTGVPVAQPAKSWDPNQLNAHPLEPVLRLGLKTAEGALRPPPGKG 540
Qy 541 S 541
Db 541 S 541

RESULT 3
US-10-615-659-13
; Sequence 13, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-13

Query Match 99.6%; Score 2843; DB 4; Length 541;
Best Local Similarity 99.6%; Pred. No. 2.5e-228;
Matches 539; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 MASSILKWVSHQSCSRSSKPRDQREAEAGSSDLSSRQDAENAEAKLGLPGQLVDIAC 60
Db 1 MASSILKWVSHQSCSRSSKPRDQREAEAGSSDLSSRQDAENAEAKLGLPGQLVDIAC 60
Qy 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQQYYSLVHGDAFISNSRNYFSQCOAL 120
Db 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQQYYSLVHGDAFISNSRNYFSQCOAL 120
Qy 121 LNRTSVNPTDIDGLRNIWIIPAAKSRGRDIVCMRDVEEILELAADHPLSRDNKVV 180
Db 121 LNRTSVNPTDIDGLRNIWIIPAAKSRGRDIVCMRDVEEILELAADHPLSRDNKVV 180
Qy 181 QKYIETPLLCIDTKFDIRQWFLVTDWNPFTIWFYKESYLPSTQRFSLDKLDSAIHLCNN 240
Db 181 QKYIETPLLCIDTKFDIRQWFLVTDWNPFTIWFYKESYLPSTQRFSLDKLDSAIHLCNN 240
Qy 241 AVQKYLKNDVGRSPLLAHNNMTSTRFQEYLQROGRGAVGWSVIYPSMKKAIHAAMKVAQ 300
Db 241 AVQKYLKNDVGRSPLLAHNNMTSTRFQEYLQROGRGAVGWSVIYPSMKKAIHAAMKVAQ 300
Qy 301 DHVEPRKNSFELYGADFVLGRDFRPWLIEINSSPTMHPSTPTVAQLCAQVOEDTIKVAVD 360
Db 301 DHVEPRKNSFELYGADFVLGRDFRPWLIEINSSPTMHPSTPTVAQLCAQVOEDTIKVAVD 360
Qy 361 RSCDIGNFELLWQPVVPPFSGSDLCVAGSVRRARROVLPVCNLKASASLLDAQPLK 420
Db 361 RSCDIGNFELLWQPVVPPFSGSDLCVAGSVRRARROVLPVCNLKASASLLDAQPLK 420
Qy 421 ARGPSAMPDPAQPPSPALQDRLGLKEEKGPLALLAPLRGAESGAAQPTRTKAAGKV 480
Db 421 ARGPSAMPDPAQPPSPALQDRLGLKEEKGPLALLAPLRGAESGAAQPTRTKAAGKV 480
Qy 481 ELPACPCRHVDSQAPNTGVPVAQPAKSWDPNQLNAHPLEPVLRLGLKTAEGALRPPPGKG 540
Db 481 ELPACPCRHVDSQAPNTGVPVAQPAKSWDPNQLNAHPLEPVLRLGLKTAEGALRPPPGKG 540
Qy 541 S 541

Db 541 S 541

RESULT 4

US-10-635-977-13

; Sequence 13, Application US/10635977
; Publication No. US20040171131A1

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN

; FILE REFERENCE: D0283A CIP

; CURRENT APPLICATION NUMBER: US/10/635,977

; PRIOR FILING DATE: 2003-08-07

; PRIOR APPLICATION NUMBER: U.S. 60/394,725

; PRIOR FILING DATE: 2002-07-09

; PRIOR APPLICATION NUMBER: U.S. 10/615,659

; NUMBER OF SEQ ID NOS: 103

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 13

; LENGTH: 541

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-635-977-13

Query Match

Best Local Similarity 99.6%; Score 2843; DB 4; Length 541;

Matches 539; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MASSILKVVVSHQSCRSRSPKQREAGSDLSRRQDAENAEAKRLGLPGQLVDIAC 60

Db 1 MASSILKVVVSHQSCRSRSPKQREAGSDLSRRQDAENAEAKRLGLPGQLVDIAC 60

QY 61 KVCQAYLGQLEHEDITSADAVEDLTEAWEDLTQYYSLVHGDAFISNSRNYFSQCQAL 120

Db 61 KVCQAYLGQLEHEDITSADAVEDLTEAWEDLTQYYSLVHGDAFISNSRNYFSQCQAL 120

QY 121 LNRITSVNPQTDIDGLRNIWIKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVV 180

Db 121 LNRITSVNPQTDIDGLRNIWIKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVV 180

QY 181 QKYIETPLLICDTKFDIRQWFLVTDWNPFTIWFYKESYLRFTQRFSLDKLDSAIHLCNN 240

Db 181 QKYIETPLLICDTKFDIRQWFLVTDWNPFTIWFYKESYLRFTQRFSLDKLDSAIHLCNN 240

QY 241 AVQKYLKNDVGRSPLLPANHMTSTRFQYLRQGRGAVGWSVIYPSMKKAIAHAMKVAQ 300

Db 241 AVQKYLKNDVGRSPLLPANHMTSTRFQYLRQGRGAVGWSVIYPSMKKAIAHAMKVAQ 300

QY 301 DHVEPRKNSFELYGADFVLGRDRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVD 360

Db 301 DHVEPRKNSFELYGADFVLGRDRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVD 360

QY 361 RSCDIGNFELLWRQVPVPPPPSGSDLCVAGSVRRARRQVLPVNLKASALLDAQPLK 420

Db 361 RSCDIGNFELLWRQVPVPPPPSGSDLCVAGSVRRARRQVLPVNLKASALLDAQPLK 420

QY 421 ARGPSAMPDPAQGPSALORDLGLKEKGLPLALLAPLRGAESGGAAQPTRTKAAKV 480

Db 421 ARGPSAMPDPAQGPSALORDLGLKEKGLPLALLAPLRGAESGGAAQPTRTKAAKV 480

QY 481 ELPACPCRHRVDSQAPNTGVPVQAQPAKSWDPNQLNAHPLEPVLRGLKTAEGALRPPPGKG 540

Db 481 ELPACPCRHRVDSQAPNTGVPVQAQPAKSWDPNQLNAHPLEPVLRGLKTAEGALRPPPGKG 540

QY 541 S 541

Db 541 S 541

RESULT 5

US-10-615-659-14

; Sequence 14, Application US/10615659

; Publication No. US20040157234A1

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN

; FILE REFERENCE: D0283 NP

; CURRENT APPLICATION NUMBER: US/10/615,659

; PRIOR FILING DATE: 2003-07-09

; PRIOR APPLICATION NUMBER: U.S. 60/394,725

; NUMBER OF SEQ ID NOS: 102

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 14

; LENGTH: 293

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-615-659-14

Query Match

Best Local Similarity 100.0%; Score 1560; DB 4; Length 293;

Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 EDITSADAVEDLTEAWEDLTQYYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 132

Db 1 EDITSADAVEDLTEAWEDLTQYYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 60

QY 133 IDGLRNIWIKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKYIETPLLICD 192

Db 61 IDGLRNIWIKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKYIETPLLICD 120

QY 193 TKFDIRQWFLVTDWNPFTIWFYKESYLRFTQRFSLDKLDSAIHLCNNAVQKYLKNDVGR 252

Db 121 TKFDIRQWFLVTDWNPFTIWFYKESYLRFTQRFSLDKLDSAIHLCNNAVQKYLKNDVGR 180

QY 253 SPLLPANHMTSTRFQYLRQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312

Db 181 SPLLPANHMTSTRFQYLRQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 240

QY 313 YGADFVLGRDRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDRSCDI 365

Db 241 YGADFVLGRDRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDRSCDI 293

RESULT 6

US-10-635-977-14

; Sequence 14, Application US/10635977

; Publication No. US20040171131A1

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN

; FILE REFERENCE: D0283A CIP

; CURRENT APPLICATION NUMBER: US/10/635,977

; PRIOR FILING DATE: 2003-08-07

; PRIOR APPLICATION NUMBER: U.S. 60/394,725

; PRIOR FILING DATE: 2002-07-09

; PRIOR APPLICATION NUMBER: U.S. 10/615,659

; NUMBER OF SEQ ID NOS: 103

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 14

; LENGTH: 293

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-635-977-14

Query Match

Best Local Similarity 100.0%; Score 1560; DB 4; Length 293;

Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 EDITSADAVEDLTEAWEDLTQYYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 132

```

Db 1 EDIDTSADAVEDTEAEWEDLTQYYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 60
QY 133 IDGLRNIIWKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKYIETPLLICD 192
Db 61 IDGLRNIIWKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKYIETPLLICD 120
QY 193 TKFDIRQWFLVTDWNPNTIWFYKESYLRSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 252
Db 121 TKFDIRQWFLVTDWNPNTIWFYKESYLRSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 180
QY 253 SP LLPAHNMWMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
Db 181 SP LLPAHNMWMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 240
QY 313 YGADFVLGRDRFRLWLEINSPTMHPSTPVTQAQLCAQVQEDTIKVAVDSCDI 365
Db 241 YGADFVLGRDRFRLWLEINSPTMHPSTPVTQAQLCAQVQEDTIKVAVDSCDI 293

RESULT 7
US-10-615-659-4
; Sequence 4, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-4
```

```

Query Match 45.5%; Score 1299; DB 4; Length 242;
Best Local Similarity 100.0%; Pred. No. 7e-100;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 IDGLRNIIWKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKYIETPLLICD 192
Db 1 IDGLRNIIWKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKYIETPLLICD 60

QY 193 TKFDIRQWFLVTDWNPNTIWFYKESYLRSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 252
Db 61 TKFDIRQWFLVTDWNPNTIWFYKESYLRSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 120

QY 253 SP LLPAHNMWMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
Db 121 SP LLPAHNMWMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180

QY 313 YGADFVLGRDRFRLWLEINSPTMHPSTPVTQAQLCAQVQEDTIKVAVDSCDI 372
Db 181 YGADFVLGRDRFRLWLEINSPTMHPSTPVTQAQLCAQVQEDTIKVAVDSCDI 240

QY 373 RQ 374
Db 241 RQ 242

RESULT 8
US-10-635-977-4
; Sequence 4, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
```

```

; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-4

Query Match 45.5%; Score 1299; DB 4; Length 242;
Best Local Similarity 100.0%; Pred. No. 7e-100;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 IDGLRNIIWKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKYIETPLLICD 192
Db 1 IDGLRNIIWKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKYIETPLLICD 60

QY 193 TKFDIRQWFLVTDWNPNTIWFYKESYLRSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 252
Db 61 TKFDIRQWFLVTDWNPNTIWFYKESYLRSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 120

QY 253 SP LLPAHNMWMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
Db 121 SP LLPAHNMWMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180

QY 313 YGADFVLGRDRFRLWLEINSPTMHPSTPVTQAQLCAQVQEDTIKVAVDSCDI 372
Db 181 YGADFVLGRDRFRLWLEINSPTMHPSTPVTQAQLCAQVQEDTIKVAVDSCDI 240

QY 373 RQ 374
Db 241 RQ 242

RESULT 9
US-10-275-595A-5
; Sequence 5, Application US/10275595A
; Publication No. US20040078804A1
; GENERAL INFORMATION:
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: LAU, Preeti
; APPLICANT: YAO, Monique G.
; APPLICANT: BURFORD, Neil
; APPLICANT: KEARNEY, Liam
; APPLICANT: POLICKY, Jennifer L.
; TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS
; FILE REFERENCE: PE-0772 USN
; CURRENT APPLICATION NUMBER: US/10/275,595A
; PRIOR FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/201,960
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/202,729
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 60/209,705
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 60/210,149
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: US 60/213,215
```


Db 114 NPVWMDGKWKVQKTYIERPLLIIFTGKFDLRQWFLVTDWNPPLTWFFYRDSYIRFSTQPFSL 173
QY 229 DKLDSAIHCNNNAVQKYLKNDVGRSPLLPAHNNWTSRFOEYLRQGRGAVNGSVIYPSM 288
Db 174 KNLDSVHLCCNSIQHLENSCHRRHPLLPDNNMWSQRFQAHLEMGAPNASTIIVPGM 233
QY 289 KKAIAHAMKVAQDHVPRKNSFELYGADFVLGRDRPWLIEINSSPTMHPSTPVTQAOLCA 348
Db 234 KDAVIHALQTSQDTVQCRKASFELYGADFVFGDFQFWLIEINASTMAPSTAVTARLCA 293
QY 349 QVQEDTIKAV-----DRSCDIGNFELLWRQPVVEPPFPFSGSDLCVAGSVRR 396
Db 294 GVQADTLRVVIDRXLDRNCDTGAFELIYKQPAVEVQYVGIRLLVBSFTIKK 345
RESULT 12
US-10-615-659-7
; Sequence 7, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 7
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-7

Query Match 33.5%; Score 956.5; DB 4; Length 352;
Best Local Similarity 54.2%; Pred. No. 4.8e-71;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;
QY 74 DIDSADAVEDLTAEWEDLTQOYVSLVHGDAFISNRNYSFQOALLNRITSVNPQTDI 133
Db 2 DIDKDLAPLYLTPEGNSLFQRYQVHGAELRHLDTVQRCEDILQOLQAVVPQIDM 61
QY 134 DGLRNIWIKPAKSRGRDIVCMDRVVEILELAADHPLSRDNKNVQVQKIETPLICDT 193
Db 62 EGRDNIWIKPGAKSRGRGIMCMHLEMLKLVNGPNVVMKDGKVVQKIERPLLIIFT 121
QY 194 KFDIRQWFLVDWNPPLTWFFYRDSYIRFSTQPFSLKNDNSVHLCCNSIQHLENSCHRH 253
Db 122 KFDLRQWFLVDWNPPLTWFFYRDSYIRFSTQPFSLKNDNSVHLCCNSIQHLENSCHRH 181
QY 254 PLLPAHNNWTSRFOEYLRQGRGAVNGSVIYPSMKAIAHAMKVAQDHVPRKNSPELY 313
Db 182 PLLPDDNNMWSQRFQAHLEMGAPNASTIIVPGMKDAVIHALQTSQDTVQCRKASFELY 241
QY 314 GADFVLGRDRFPWLIEINSSPTMHPSTPVTQAOLCAQVQEDTIKAV-----DRSCDIGNFE 369
Db 242 GADFVFGEDFQFWLIEINASTMAPSTAVTARLCAGVQADTLRVVIDRMLDRNCDTGAFE 301
QY 370 LLMRQPVVEPPFPFSGSDLCVAGV-SVRRAR 398
Db 302 LIYQPVTTSPASTPRPSCLLPMYSSTRAR 331

RESULT 13
US-10-635-977-7
; Sequence 7, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 7
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-7

; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 7
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-7
Query Match 33.5%; Score 956.5; DB 4; Length 352;
Best Local Similarity 54.2%; Pred. No. 4.8e-71;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;
QY 74 DIDSADAVEDLTAEWEDLTQOYVSLVHGDAFISNRNYSFQOALLNRITSVNPQTDI 133
Db 2 DIDKDLAPLYLTPEGNSLFQRYQVHGAELRHLDTVQRCEDILQOLQAVVPQIDM 61
QY 134 DGLRNIWIKPAKSRGRDIVCMDRVVEILELAADHPLSRDNKNVQVQKIETPLICDT 193
Db 62 EGRDNIWIKPGAKSRGRGIMCMHLEMLKLVNGPNVVMKDGKVVQKIERPLLIIFT 121
QY 194 KFDIRQWFLVDWNPPLTWFFYRDSYIRFSTQPFSLKNDNSVHLCCNSIQHLENSCHRH 253
Db 122 KFDLRQWFLVDWNPPLTWFFYRDSYIRFSTQPFSLKNDNSVHLCCNSIQHLENSCHRH 181
QY 254 PLLPAHNNWTSRFOEYLRQGRGAVNGSVIYPSMKAIAHAMKVAQDHVPRKNSPELY 313
Db 182 PLLPDDNNMWSQRFQAHLEMGAPNASTIIVPGMKDAVIHALQTSQDTVQCRKASFELY 241
QY 314 GADFVLGRDRFPWLIEINSSPTMHPSTPVTQAOLCAQVQEDTIKAV-----DRSCDIGNFE 369
Db 242 GADFVFGEDFQFWLIEINASTMAPSTAVTARLCAGVQADTLRVVIDRMLDRNCDTGAFE 301
QY 370 LLMRQPVVEPPFPFSGSDLCVAGV-SVRRAR 398
Db 302 LIYQPVTTSPASTPRPSCLLPMYSSTRAR 331

RESULT 14
US-10-756-149-5267
; Sequence 5267, Application US/10756149
; Publication No. US20050181375A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ziolknik, Albert
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER. COMPOSITIONS AN
; TITLE OF INVENTION: METHODS OF SCREENING FOR MODULATORS OF METASTATIC CANCER
; FILE REFERENCE: file
; CURRENT APPLICATION NUMBER: US/10/756,149
; CURRENT FILING DATE: 2004-01-12
; NUMBER OF SEQ ID NOS: 5818
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 5267
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-756-149-5267

Query Match 33.5%; Score 956.5; DB 5; Length 352;
Best Local Similarity 54.2%; Pred. No. 4.8e-71;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;
QY 74 DIDSADAVEDLTAEWEDLTQOYVSLVHGDAFISNRNYSFQOALLNRITSVNPQTDI 133
Db 2 DIDKDLAPLYLTPEGNSLFQRYQVHGAELRHLDTVQRCEDILQOLQAVVPQIDM 61
QY 134 DGLRNIWIKPAKSRGRDIVCMDRVVEILELAADHPLSRDNKNVQVQKIETPLICDT 193

```
Db 62 EGDRIWIIVKPAKSGRIGMCMHLEMLKLVNGNPVVMKDGKVVQYIERPLLIFFT 121
Qy 194 KFDIRQWFLVTDWNPITWIFWYKESYLRFSYRFSQFSLDKLDSAIHLCCNNAVQKYLKNDVGRS 253
Db 122 KFDLRQWFLVTDWNPITWIFWYRDSYIRFSTQFSLKNDNSVHLCCNNSIQKHLNSCHRH 181
Qy 254 PLLPAHNNWTSTRFOEYIQRQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFELY 313
Db 182 PLLPPDNMWSQRFQAHLEQMGAPNAWSTIIVPGMKDAVIHALQTSQDTVQCRKASFELY 241
Qy 314 GADFLVGRDPRPWLIEINSPTHPSTPTVTAQLCAQOQEDTIKAV-DRSCDIGNFE 369
Db 242 GADFLVGRDPRPWLIEINSPTHPSTPTVTAQLCAQOQEDTIKAV-DRSCDIGNFE 369
Qy 370 LLWRQVPPPPFSGSDLCVAGV-SVRRAR 398
Db 302 LIYKQPVTTSPASTPRPSCLLPMYSDTRAR 331
```

RESULT 15

```
US-10-108-260A-4209
; Sequence 4209, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1 full length cdna
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 4209
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-4209
```

```
Query Match 33.1%; Score 944.5; DB 4; Length 326;
Best Local Similarity 54.3%; Pred. No. 4.3e-70;
Matches 176; Conservative 55; Mismatches 86; Indels 7; Gaps 2;

Qy 74 DIDTSADAVEDLTEAEWEDLTQYYSLVHGDAFISNSRNYFSQCOALLNRITSVNPQTDI 133
Db 2 DIDKLEAPLYLTPEGWSLFLQYYQVVEGAEHLRLDTQVQRCEDILQQLQAVVPQIDM 61
Qy 134 DGLRNWIIKPAKSGRDIIVCMRVEIILEAAADHPLSRDKNKVVQYIETPLIICDT 193
Db 62 EGDRIWIIVKPAKSGRIGMCMHLEMLKLVNGNPVVMKDGKVVQYIERPLLIFFT 121
Qy 194 KFDIRQWFLVTDWNPITWIFWYKESYLRFSYRFSQFSLDKLDSAIHLCCNNAVQKYLKNDVGRS 253
Db 122 KFDLRQWFLVTDWNPITWIFWYRDSYIRFSTQFSLKNDNSVHLCCNNSIQKHLNSCHRH 181
Qy 254 PLLPAHNNWTSTRFOEYIQRQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFELY 313
Db 182 PLLPPDNMWSQRFQAHLEQMGAPNAWSTIIVPGMKDAVIHALQTSQDTVQCRKASFELY 241
Qy 314 GADFLVGRDPRPWLIEINSPTHPSTPTVTAQLCAQOQEDTIKAV-DRSCDIGNFE 369
Db 242 GADFLVGRDPRPWLIEINSPTHPSTPTVTAQLCAQOQEDTIKAV-DRSCDIGNFE 369
Qy 370 LLWRQVPPPPFSGSDLCVA 390
Db 302 LIYKQGPAPNMQVSPERNAPICPA 325
```

Search completed: April 4, 2006, 12:57:59
Job time : 78.9322 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:46:06 ; Search time 21.1171 Seconds
(without alignments)
2118.072 Million cell updates/sec

Title: US-10-635-977-2
Perfect score: 2854
Sequence: 1 MASSILKVVVSHQSCRSR.....LRGLKTAEALRPPPGKGS 541

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA*
1: /cgn2_6/ptodata/1/iaa/5 COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/6 COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/H COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/pCTUS COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/RE COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	280.5	9.8	439	2	US-10-104-047-3658
2	271.5	9.5	592	2	US-10-104-047-3371
3	212	7.4	561	2	US-09-270-767-46703
4	183	6.4	404	2	US-10-104-047-3621
5	154.5	5.4	260	2	US-09-270-767-62411
6	154.5	5.4	507	2	US-09-270-767-46787
7	151	5.3	259	2	US-09-248-796A-14603
8	144.5	5.1	753	2	US-09-538-092-31
9	126	4.4	305	2	US-09-270-767-46577
10	126	4.4	2482	2	US-09-252-991A-16967
11	125.5	4.4	802	2	US-09-823-240A-2
12	123	4.3	884	6	5208144-8
13	120	4.2	542	2	US-09-949-016-6778
14	120	4.2	554	2	US-09-949-016-11687
15	120	4.2	950	2	US-09-252-991A-25927
16	114.5	4.0	1319	1	US-08-290-731C-2
17	114.5	4.0	1336	1	US-08-290-731C-6
18	113.5	4.0	2074	2	US-09-491-356C-9
19	111	3.9	806	2	US-09-902-540-14992
20	110.5	3.9	560	2	US-09-252-991A-25999
21	107.5	3.8	535	2	US-09-252-991A-17140
22	107	3.7	419	2	US-10-002-344A-237
23	107	3.7	1183	2	US-09-134-001C-3530
24	106.5	3.7	328	2	US-09-252-991A-17729
25	106.5	3.7	574	2	US-09-270-767-45812
26	106	3.7	679	2	US-09-252-991A-18857
27	106	3.7	780	2	US-09-252-991A-32892

28	105.5	3.7	542	2	US-09-252-991A-32139	Sequence 32139, A
29	105.5	3.7	1706	2	US-09-252-991A-31760	Sequence 31760, A
30	105	3.7	416	2	US-09-252-991A-32660	Sequence 32660, A
31	105	3.7	561	2	US-09-252-991A-23080	Sequence 23080, A
32	105	3.7	781	2	US-09-949-016-9170	Sequence 9170, Ap
33	104.5	3.7	582	2	US-08-906-865-3	Sequence 3, Appl1
34	104.5	3.7	582	2	US-09-129-668-3	Sequence 18696, A
35	104	3.6	406	2	US-09-252-991A-18696	Sequence 8369, Ap
36	104	3.6	955	2	US-09-949-016-8369	Sequence 21872, A
37	103.5	3.6	248	2	US-09-252-991A-21872	Sequence 8, Appl1
38	103.5	3.6	798	2	US-09-150-460B-8	Sequence 7, Appl1
39	103.5	3.6	875	2	US-09-150-460B-7	Sequence 20033, A
40	103.5	3.6	1290	2	US-09-150-460B-6	Sequence 32630, A
41	103	3.6	420	2	US-09-252-991A-20033	Sequence 159, App
42	103	3.6	515	2	US-09-252-991A-32630	Sequence 3402, Ap
43	103	3.6	691	2	US-09-758-759-159	Sequence 12305, A
44	103	3.6	710	2	US-10-104-047-3402	
45	102.5	3.6	398	2	US-09-902-540-12305	

ALIGNMENTS

RESULT 1

US-10-104-047-3658
; Sequence 3658, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cdna
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 3658
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3658

Query Match	9.8%	Score 280.5;	DB 2;	Length 439;
Best Local Similarity	25.0%	Pred. No. 3e-18;		
Matches	101;	Conservative	65;	Mismatches 135; Indels 103; Gaps 18;
QY	82	VEDTEARWE	-----DLT-----	QOYSLVHGDAPFISNRNYFSQC 117
DB	65	VEVKDEGEWDFYWCDSWLRNFHDHYMDEHVRISHFRNHVELTRKNYVVKLFRKQL 124		
QY	118	QALLNRITSVN	---PQT-----	DIQGLN---IWIKPAKSRGRDIVCWDVRE 160
DB	125	EREAGLAAKCDFFPKTFEMPCYHLEFVEFRKNPGITWIMKPVARSQKGIFLFRRLK 184		
QY	161	EILELAADHPLSRDNK	-----WVQKYLETPLLCDTKFDIROWFLVTDWNPITWF 213	
DB	185	DIVDM-RKDTSSDDQDDIPVENYVAQRYLENFYLCGRKFDLEVYLVMS	-----V 236	
QY	214	YKESYLRPSTQFSLDKLDSAIHLCNNAVKYLKNDVGRSPLLPANH	-----MMTSTRFQ 268	
DB	237	FAECLLWSGHR	-----QDVHLTNVAVQK-----	TSPDYHPKKGCKTQLQFR 279
QY	269	EYL-QROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFELYGADFLGRDRFWL 327		
DB	280	QYLASKHGPEAV	--ETLFRDIDNIFVKSLSQVQVITISDKHCFELYGYDILIDQDLKMWL 337	
QY	328	IEINSSPTMHPSTPVTALCAQVOEDITKVA	-----VDRSCDIGNFELLWRQPVVE--- 378	
DB	338	LEVNASPLTASSQEDVELKTCILEDLHVVDMEARLTGCRKRVGGFFLMMNDGVPVREE 397		
QY	379	PPFPFSGDLCAVGVS	-----RRARRQVLPCNL-----KASA 411	

Db 398 GAPDLSGNGFVNTHTLGCYNDKRLQRLF--CSLQVKKASS 439

RESULT 2

US-10-104-047-3371
; Sequence 3371, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3371
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3371

Query Match 9.5%; Score 271.5; DB 2; Length 592;
Best Local Similarity 29.8%; Pred. No. 3.6e-17;
Matches 72; Conservative 43; Mismatches 80; Indels 47; Gaps 8;
Qy 113 YFQSQALLNRITSNVNPOTDIDGLRNIIIPKAKSRGRDIVCMRVVEIELEAAADHPL 172
Db 195 YFQERQMLGK-----HSYWKCPAELSRGGLIFSDPKFI----- 232
Qy 173 SRNKVVQKYIETPLICTKDIQWFLVTDWNPITWIFYKESYLRFSQFSLDKL- 231
Db 233 -FDMYIVQKYSINPLIGRYKCDLRIYVCVTGFKPLTIYVQSLVRFATEKFDLSNLQ 291
Qy 232 DSATHLCNNAVQK-----VLKNDVGRSPLLPAHNMWTSRFOYLQROGRGAVW---GS 282
Db 292 NNVAHLNNTSINKSGASYEKIKEVIGHG-----CKWTLRFRFSYLRK-----WDVDDL 339
Qy 283 VIYPSMKKAIAHAMKVAQDHVEPRKNSFELYGADFVLGRDFRPWLIBINSPP--TMHPST 340
Db 340 LLWKIHRMWLITLTAIPSPVFAANCFFELFGFDLILDDNKLKPLLEVNYSPALTLDCST 399
Qy 341 PV 342
Db 400 DV 401

RESULT 3

US-09-270-767-46703
; Sequence 46703, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46703
; LENGTH: 561
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-46703

Query Match 7.4%; Score 212; DB 2; Length 561;
Best Local Similarity 29.2%; Pred. No. 1.9e-11;
Matches 57; Conservative 37; Mismatches 75; Indels 26; Gaps 6;
Qy 179 VWQKYIETPLICTKDIQWFLVTDWNPITWIFYKESYLRFSQFSLDKL-DSAI 235
Db 1 VVSKYIIVDPLCIDGHKCDLRYVVLVTSFDPLIILYBEGIVRLATVKYDRHADNLNPPCM 60

Qy 236 HLCNNAVQKYLKN-----DVGRSPLLPAHNMWTSRFOYLQROG---RGAYWG- 281
Db 61 HLCNYSINKYHSDYIRSSDAQDEVG-----HKWTLRALLRHLKQLQSCDTRQLMINI 112
Qy 282 -SVIYPSMKKAIAHAMKVAQDHVEPRKNSFELYGADFVLGRDFRPWLIBINSPTMHPST 340
Db 113 EDLIIKAVLACQSIISACRMFVPGNNGCFELYGFDLILIDNAHKPWILLEINLSPSMGVDS 172
Qy 341 PVTAQLCAQVQEDTI 355
Db 173 PLDTKVKSCIMADLL 187
RESULT 4
US-10-104-047-3621
; Sequence 3621, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3621
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3621

Query Match 6.4%; Score 183; DB 2; Length 404;
Best Local Similarity 25.5%; Pred. No. 7.3e-09;
Matches 82; Conservative 38; Mismatches 118; Indels 84; Gaps 17;
Qy 6 LKWVSHQSCSRSRKPRQOREEAGSSDLSSRQDAENAEAK-----LRGLPGQLVDIA 59
Db 119 LKW-----CEVKSRSYSGSFRE--GEQLLYQLPNNKLLTTKIGLLSTLRG-RARAMSRA 169
Qy 60 CKVQAVYLGLEHEDIDTSADAVEDLTEAEWEDLTQYYSLVHGDAFISNRNYFSQCA 119
Db 170 SKVPGGVQARLEK---DAAAPALEDL---PW-----TSPGYLRQVR 205
Qy 120 LLNRITSVNPQT---DIDGLR-----NIWIKPAKSRGRDIVCMRVVEIELELA 166
Db 206 L--RMBEFPETVRLDLKHEREAFFLTFDETQIWKCTASNOGKGFLLRNQBEVAALQ 263
Qy 167 A-----ADHPLSRDNKW-----VQKYIETPLICTKDIQWFLVTDWNPITWIFYK 215
Db 264 AKTRSMEDDPIHKTTPRGPQARVVQYIQNPLLVGKRKFDVRSYLIACCTPYMI-FFG 322
Qy 216 ESYLRFSQFSLDKLSDSAIHLNNAVQKYLKNDVGRSP---LLPAHNMWTSRFOEYL- 271
Db 323 HGVARLUTLSLYDPHSSDLGGHLTNQFMQK-----KSPLYMLLKERTVMSMEHLNRYIS 375
Qy 272 ---QROGRGAVMGSVIYPSMK 289
Db 376 DTFWKARGLAKDW---VFTTLK 394
RESULT 5
US-09-270-767-62411
; Sequence 62411, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17


```
QY 278 AVWGSVIYPSMKKAIHAHAKVADHVEPRKNSPELYGADFLVGRDPRFWLIEINSSPTMH 337
Db 643 NI-KEQHSITNDVFLAANNVNRKLNFOPLNAPETYGVDPLIDNSYEVKLEINAPDPK 701
QY 338 PSTPVTQAQCAQVQEDTIKAV 359
Db 702 QTKGDKLNLIDELFDDTVKVCV 723

RESULT 9
US-09-270-767-46577
; Sequence 46577, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 46577
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; US-09-270-767-46577

Query Match 4.4%; Score 126; DB 2; Length 305;
Best Local Similarity 23.6%; Pred. No. 0.0015;
Matches 42; Conservative 36; Mismatches 62; Indels 38; Gaps 8;

QY 180 VQKIETPLICTKEDIROFWLTDWNPPLTIWYFYESYL-RFSTORFSLDKLDSAIHLC 238
Db 1 VQSFVORPYLDVGHKFDIGYVVVITSVNPRLRYITGDVLFRCYCPVKYHPFDAEN----- 55
QY 239 NNAVQYKLVKNDVGRSPLPA-----HNW---TSTRFOYVLOQRG--GAVWGSVI 284
Db 56 ---VDKIYVGD---DYLPTWEVPSLRKYNRFGGSMRTVFEAYVRDQGDPAQIWPQVE 108
QY 285 Y-----PSMKKAIHAHAKVADHVEPRKNSPELYGADFLVGRDPRFWLIEINSSPTM 336
Db 109 HIVRTTIAKEKDVILNLSYRTH-----NFFDLMRFDLFIDEDLKVFLMEANWSPNL 161

RESULT 10
US-09-252-991A-16967
; Sequence 16967, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16967
; LENGTH: 2482
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (1766)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
; US-09-252-991A-16967

Query Match 4.4%; Score 126; DB 2; Length 2482;
Best Local Similarity 19.7%; Pred. No. 0.044;
Matches 122; Conservative 79; Mismatches 207; Indels 210; Gaps 29;
```

```
QY 71 EHEDIDT-----SADAVEDLTAEWEDLTQQVYVLVHGDAFISNSRNFYSQCOALLN 122
Db 308 DHETVDAERARLAGPRDAMRSVVGALCEELVRIKDSL---DLFVSRDRGHSELDALLA 364
QY 123 RITSVNPQTDIDG-----LNNIWIIPAAKSRGR--DIVCMRVEEILEAA----- 167
Db 365 PLKQIADTLAVLGFQGPVKVILQDLQDVIHALAAGRRPSDAIILMDVAGALLYVEATLAGM 424
QY 168 ---ADHPLSRDNKWWVQKIET-PLLICDTKFDIRO-----WFLVTDWNPPLTIWYKE- 216
Db 425 AGPGDERNSEESRLPTTVDVAIQHQLVIKEARNGLEQAKDAIEFFIASQNMHEHLARVPEL 484
QY 217 -SVLRESTQRFSLDKLDSALHLCNNAVQYKLVKNDVGRSPLLPANNWTS-----TRF 267
Db 485 LTQVRGGLAMIPLERAATLLETCTNRYIQEQL---LARK---AVPDWQSLDITLADAITSV 537
QY 268 QEYLQQRGAVGWSVIYPSMKKAIHAHAKVADHVEPRKNSPELYGADFLVGRDPRFWL 327
Db 538 EYTLERL-----SEDHASQSDLI---LDVAEDSLAN-----LGYTLKP-- 572
QY 328 IEINS-----SPTMHPSTPVTQOLCAQ-----VOEDTIKVA 358
Db 573 --NSSAPAEPLSGSPAAIESPAAEPEPEAVVEVAETAEPQPPADTAPAEAREDAPOLA 629
QY 359 VDSCDIG-----NFELLWRQPV---VEPPP-----F 382
Db 630 SDNWTTLGEVADPAGPEPSLDLALDLPDSDSAEVPFALPEVVEESGQPSQTPAPARSLDDF 689
QY 383 SGSDLCVAGSVRRARRRQVLVPCNLKASALLDAQPLKARGPSAMPD---PAQ----- 432
Db 690 SLDEIDLSGLD-----LPA-----DAAP--ASGPAALADWSLPEQWGLGDDL 729
QY 433 -GPSPALORDLGLKE-----EKGLPLALLAPLRGAESG---GAQP-----TRT 474
Db 730 AQPTQAGETLIDLSLEBPALSFDAPLESLEPLPALEPFDGSAEQELVLDLDDLPPLDVALP 789
QY 475 KAAGKV-----ELPACPCRHVDQAPNTGVPVQAQPAKSWDPNQLNAHPL 518
Db 790 EAEGEVSANEGSLEELDLSDDLPEVQLPEAEAEAPPAEAALASAPALSLEAVMAAPV 849
QY 519 EPVLRLGLKTAEGALRPPP 536
Db 850 QPINPPAQNPVSLPPPP 867

RESULT 11
US-09-823-240A-2
; Sequence 2, Application US/09823240A
; Patent No. 6716597
; GENERAL INFORMATION:
; APPLICANT: Frank B. Gertler
; APPLICANT: James E. Bear
; APPLICANT: Jurgen Wehland
; APPLICANT: Joseph Loureiro
; TITLE OF INVENTION: Methods and Products for Regulating Cell
; TITLE OF INVENTION: Motility
; FILE REFERENCE: M00656.70064.US
; CURRENT APPLICATION NUMBER: US/09/823,240A
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/194,564
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-823-240A-2

Query Match 4.4%; Score 125.5; DB 2; Length 802;
Best Local Similarity 24.7%; Pred. No. 0.0081;
Matches 78; Conservative 15; Mismatches 110; Indels 113; Gaps 15;
```



```
; Sequence 11687, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11687
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Human
; ORGANISM: Human
US-09-949-016-11687
```

```
Query Match 4.2%; Score 120; DB 2; Length 554;
Best Local Similarity 22.5%; Pred. No. 0.015;
Matches 75; Conservative 38; Mismatches 124; Indels 96; Gaps 18;

QY 254 PLLPAHMMTSTRFOEYLQKRGAGVSGSVIYPSMKKAIHAAMKVAQDHVEPRKNSFELY 313
DB 245 PLLPGNDV-----GRSS-YGAM---QVQVFDYAVIVLSHAVSPILARSYPNR 287
QY 314 GADFEVLGR-----DFRPLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVDKSCD 364
DB 288 DAESTLGRLLIKVTQEVIDYRWIKKWSKA-HPSPGMDSRI--KIKE---RIATCNGEQ 341
QY 365 IGFNELLWRQPVVEPP-----PFGSDLCVAGSVRRARR-----QVLPVCNLKA- 409
DB 342 TQN-----REF--ESPYGQRLTSLSPQLLSGSSASSVSSLSGSDVSDTTPCTTPSV 394
QY 410 -SASLLDAQPLKARGSPAMPDPAQPPSPALQDGLKEKGLPLALLAPLRGAA-----E 464
DB 395 YQSLQAPAFMLMAGLPTALPMPS-GKQPTTSRTLIMTTNNQTRFTIPPTLGVAPVPCR 453
QY 465 SGGAAPTRTKAAGKVELPACPCRHVDSQAPNTGVPVAQPAKSWDPNQLNAHLEPVLRG 524
DB 454 QAGVEGTASLKAVHMHSSPAIP-----SASP-----PLSSP-----HLYHKQHG 494
QY 525 LK-----TAEGALRPPPGKG 540
DB 495 MKLSMKGSHGHTGGGYSSVSGGVPRPVGNRG 527
```

```
RESULT 15
US-09-252-991A-25927
; Sequence 25927, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25927
; LENGTH: 950
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25927
```

```
Query Match 4.2%; Score 120; DB 2; Length 950;
Best Local Similarity 25.1%; Pred. No. 0.036;
Matches 53; Conservative 17; Mismatches 63; Indels 78; Gaps 9;

QY 361 RSCDIGNPELLWRQPVVEPPF---SGSDLCVAGSVR-----RARRQVLPVCNLKA 409
DB 728 RTADVS-----RQPLRQPPALLQRRGDVLDASLAPRPPGFGAARARLLLSPCACRRGR 781
QY 410 SASLLDAQP-----LKARGPSAM-----PDPAQG-----PFSPALQDGLK 446
DB 782 ATGLGQRPQGRGERLSRCPGAATAAGPGGTAAHARGGGDGLRRRHHPPAPARPPGGLA 841
QY 447 EKGLPLALLAPLRGAESGGAAQPTTKAAGKVELPACPCRHVDSQAPNTGVPVAQPAK 506
DB 842 P-----APAGGAADFGAAQPGRWPA-----PAARRQR 869
QY 507 SWDPNQLNAHLEPVLRLKTAEGALRPPPG 537
DB 870 TWRTRRRS-----LRTLERPAGVPPAPP 894
```

Search completed: April 4, 2006, 12:46:55
Job time : 22.1171 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.
OM protein - protein search, using sw model
Run on: April 4, 2006, 12:43:06 ; Search time 81.9545 Seconds
(without alignments)
4657.352 Million cell updates/sec
Title: US-10-635-977-2
Perfect score: 2854
Sequence: 1 MASSILKVVVSHQSCSRSSR.....LRGLKTABGALRPPPGKGKS 541

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2166443 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 2166443
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Description	
Result No.	Score	Query Match %	Length	ID		
1	2741	96.0	747	Q5JZ84	homo sapien	Q5JZ84
2	1753.5	61.4	781	Q8CON7	mus musculus	Q8CON7
3	1726.5	60.5	518	Q8COV2	mus musculus	Q8COV2
4	1052	36.9	704	Q8BV51	mus musculus	Q8BV51
5	1043.5	36.6	744	Q9H876	homo sapien	Q9H876
6	981.5	34.4	434	Q8NDN8	homo sapien	Q8NDN8
7	957.5	33.5	352	Q6AWA3	homo sapien	Q6AWA3
8	956.5	33.5	352	1	homo sapien	1
9	941.5	33.0	534	Q4KMS8	homo sapien	Q4KMS8
10	834	29.2	266	Q4RY08	tetraodon n	Q4RY08
11	807.5	28.3	261	Q922T0	mus musculus	Q922T0
12	746.5	26.2	572	Q58CT2	bos taurus	Q58CT2
13	722.5	25.3	992	Q7Q156	anopheles g	Q7Q156
14	584.5	20.5	501	Q7PM91	anopheles g	Q7PM91
15	536.5	18.8	719	Q7VM92	anopheles g	Q7VM92
16	535.5	18.8	756	Q5BH11	anopheles g	Q5BH11
17	524.5	18.4	281	Q6ZU95	homo sapien	Q6ZU95
18	524	18.4	375	Q4V8C1	rat	Q4V8C1
19	442.5	15.5	331	Q5TN29	anopheles g	Q5TN29
20	385	13.5	101	Q96GG8	homo sapien	Q96GG8
21	345.5	12.1	461	Q641W7	rattus norv	Q641W7
22	345	12.1	464	Q9D570	mus musculus	Q9D570
23	344.5	12.1	523	Q34TUI	dictyosteli	Q34TUI
24	335	11.7	1075	Q7QT13	giardia lam	Q7QT13
25	331.5	11.6	794	Q6BFH6	paramacium	Q6BFH6
26	328	11.5	496	Q96119	anopheles g	Q96119
27	328	11.5	496	Q9VX74	anopheles g	Q9VX74
28	317	11.1	341	Q3VX47	homo sapien	Q3VX47
29	315.5	11.1	618	Q4R7H0	macaca fasc	Q4R7H0
30	314	11.0	403	2	giardia lam	2
31				Q7QZC8	giardia lam	Q7QZC8

32	311	10.9	673	2	Q6ZVT0	HUMAN	Q6ZVT0
33	307	10.8	559	2	Q8CFV5	MOUSE	Q8CFV5
34	307	10.8	600	2	Q5T2W4	HUMAN	Q5T2W4
35	307	10.8	1333	2	Q8CHB8	MOUSE	Q8CHB8
36	303.5	10.6	1048	2	Q9P1V5	HUMAN	Q9P1V5
37	303.5	10.6	1226	2	Q9UPZ4	HUMAN	Q9UPZ4
38	303.5	10.6	1277	2	Q6EMB2	HUMAN	Q6EMB2
39	303.5	10.6	1299	2	Q5R978	PONPY	Q5R978
40	301	10.5	438	2	Q8C125	MOUSE	Q8C125
41	301	10.5	1339	2	Q5F498	CHICK	Q5F498
42	299.5	10.5	423	1	TTLL1	HUMAN	TTLL1
43	298.5	10.5	423	1	TTLL1	MOUSE	TTLL1
44	298.5	10.5	423	1	Q543S4	MOUSE	Q543S4
45	298.5	10.5	423	2	Q5PPI9	RAT	Q5PPI9

ALIGNMENTS

RESULT 1
Q5JZ84 HUMAN PRELIMINARY; PRT; 747 AA.
AC Q5JZ84;
DT 10-MAY-2005 (Tremblrel. 30, Created)
DT 10-MAY-2005 (Tremblrel. 30, Last sequence update)
DE OTTHUMP0000028514 (Fragment).
GN ORFNames=RP3-355C18.2-002;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP NUCLEOTIDE SEQUENCE.
RA Cobley V.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022327; CAI42686.1; -; Genomic DNA.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_lygase.
DR Pfam; PF03133; TTL; 1.
FT NON_TER 1
SQ SEQUENCE 747 AA; 83854 MW; F90948E159BBE589 CRC64;
Query Match 96.0%; Score 2741; DB 2; Length 747;
Best Local Similarity 94.0%; Pred. No. 1.1e-181;
Matches 529; Conservative 4; Mismatches 6; Indels 24; Gaps 4;

Qy	1	MASSILKVVVSHQSCSRSSR	KPRDQREAGSSDLSSRQDAENAEAKLGLPGQLVDIAC	60
Db	187	MASSILKVVVSHQSCSRSSR	KPRDQREAGSSDLSSRQDAENAEAKLGLPGQLVDIAC	246
Qy	61	KVCQAYLGQLBHEIDITSADAVEDL	TEAWEDLTQQYYSLVH---GDAP	107
Db	247	KVCQAYLGQLBHEIDITSADAVEDL	TEAWEDLTQQYYSLVQVPLGSLCIFIQKVM	306
Qy	108	-----SNSRNVFSOQALLNRTSVN	POTDIDGLRNIIKPAKSRGR----	158
Db	307	MSFPPPTARD--ROQALLNRTSVN	POTDIDGLRNIIKPAKSRGRSPD	364
Qy	159	VEEILEAAADPHLSRDKNKVVQKYI	ETPLICTDKFIQWFLVTDNPLTIWFK	218
Db	365	VEEILEAAADPHLSRDKNKVVQKYI	ETPLICTDKFIQWFLVTDNPLTIWFK	424
Qy	219	LRFTQRFSLDKLSAHLNNAVQKYL	KNDVGRSPLLPAAHMMTSTRFQYLR	278
Db	425	LRFTQRFSLDKLSAHLNNAVQKYL	KNDVGRSPLLPAAHMMTSTRFQYLR	484
Qy	279	VMGSVIYPSMKKATAHAKVQAQDH	VEPRKNSFELYGADFVLGRDPRPWL	338
Db	485	VMGSVIYPSMKKATAHAKVQAQDH	VEPRKNSFELYGADFVLGRDPRPWL	544

```

QY 339 STPVTATLCAQOVEDTIKVAVDSSCDIGNFELLWROPVVEPPPPSGSDLCVAGSVRRAR 398
Db 545 STPVTATLCAQOVEDTIKVAVDSSCDIGNFELLWROPVVEPPPPSGSDLCVAGSVRRAR 604
QY 399 RQVLVPVNCILKASASLLDAQPLKARGPSAMPDPQAQGPSPALQRLDGLKEEKGJPLALLAP 458
Db 605 RQVLVPVNCILKASASLLDAQPLKARGPSAMPDPQAQGPSPALQRLDGLKEEKGJPLALLAP 664
QY 459 LRGAASGGGAQPTRTYKAGKVLPACPCRHVDSPAPNTGVPVQAQPAKSWDPNQLNAHPL 518
Db 665 LRGAASGGGAQPTRTYKAGKVLPACPCRHVDSPAPNTGVPVQAQPAKSWDPNQLNAHPL 724
QY 519 EPVLRGLKTAEGALRPPPGKGS 541
Db 725 EPVLRGLKTAEGALRPPPGKGS 747

RESULT 2
O8CON7_MOUSE PRELIMINARY; PRT; 781 AA.
AC Q8CON7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus adult male testis cDNA, RIKEN full-length enriched
DE library, clone:4933401B17 product:hypothetical Tubulin-tyrosine ligase
DE containing protein, full insert sequence.
DE Names:170001901Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutharia; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis; DOI=10.1016/S0076-6879(99)03004-9;
RX MEDLINE=99279233; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RN Meth. Enzymol. 303:19-44(1999).
[2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis; DOI=10.1039/35055500;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schram L.M., Staehli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Schenkel M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RN Nature 409:685-690(2001).
[3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[4]

```

```

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis; DOI=10.1101/gr.145100;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Suganara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis; DOI=10.1101/gr.152600;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W., Hirozane T.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Kasukawa T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiraiki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR ENBL; AK030151; BAC26811.1; -; mRNA.
DR Ensembl; ENSMUSG00000022388; Mus musculus.
DR MGI; MGI:1922902; 170001901Rik.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR InterPro; IPR004344; Tub_Tyr_ligase.
DR Pfam; PF03133; TTL; 1.
DR PROSITE; PS00237; G-PROTEIN_RECEP_Fl_1; UNKNOWN_1.
KW Hypothetical protein; Ligase.
SQ SEQUENCE 781 AA; 89398 MW; 9E98793C3351C3DE CRC64;
Query Match 61.4%; Score 1753.5; DB 2; Length 781;
Best Local Similarity 63.0%; Pred. No. 4.9e-113;
Matches 345; Conservative 69; Mismatches 113; Indels 21; Gaps 8;
QY 1 MASSILKXWVSHQS-CSSRSRSPRQREBAGSSDLSSRODAENAEAKRLGQLVDIA 59
Db 206 VAASILKXWVHLQNYC---SKVGKSKKEAKNSDPKDPENPDLKLSLSQVVDTA 262
QY 60 CKVCQAYLGQLEHEDITSADAVEDTEAWEDLTQYYSLVHGDAFISNRNYSFSCQA 119
Db 263 CKVCQAYLGQLEHEDIDVSEASTEALSEEWNDLTQYYLLVHGNSITDSKSYFAQCOA 322
QY 120 LLMRTISVNPQTDIGLRNIIWKPAKSGRDIIVCMRVREITLLEAAADHPLSRDNKW 179
Db 323 LLSKISSVNPQTDIGLRNIIWKPAKSGRDIIVCMRVREITLLEAAADHPLSRDNKW 382
QY 180 VQKIETPLICTKFDIROWFLVDNPLTIWFKESYLFSTQRFSLDKLSAHLN 239
Db 383 VQKIETPLIYTKFDIROWFLVDNPLTIWFKESYLFSTQRFSLDKLSAHLN 442
QY 240 NAVQVKLNDVGRSPLLPAINMMTSTRFQYLRQGRGAVGWSVIYPSMKAIHAHV 299
Db 443 NSIQRLKNDKERSPLLPCHNMMTSTRFQYLRQGRGGSIIYPSMKRAVTNAMRVA 502

```

QY 300 QHVEPRKNSFELYGADFVLGRDFRFLWLIENSSPTMHSTPTVTAQLCAQVQEDTIKAV 359
 DB 503 QDHVEARKNSFELYGADFVLGRDFRFLWLIENSSPTMHSTPTVTAQLCAQVQEDTIKVV 562
 QY 360 ----DRSCDIGNFELLWRQVVEPPFPSPGSDLCVAGSVRRARVLPVNLKASASLID 415
 DB 563 DRKLDNRCDIGNFELLWRQVPAVELPPFNGSDLCVAGSVRRARVLPVNLKASASLID 622
 QY 416 AQLKARGSPAMPDPAQGPSPALQRLDLGLKEKGLPLALLAPLRGAASGGAQAQPTRTK 475
 DB 623 APP-KVRSARALMETVIRPRTVTRQDWKREAKVLTSTWSPVMD-AEVGRGAKPIYAF 580
 QY 476 AAGKVELPACPCRHVDSQAPNTGVPVPAQPAKSWDNPQLNAHLEPVL--RGLKTABGALR 533
 DB 681 EYNDYQ-----HVDNKS HKSGYTRVQSSKV--PGVTLTSAQHPALFAQTMKPTQMTSS 731
 QY 534 PPPGGKGS 541
 DB 732 PPPTASGN 739
 RESULT 3
 Q8COV2_MOUSE
 ID Q8COV2_MOUSE PRELIMINARY; PRT; 518 AA.
 AC Q8COV2; 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930524K07 product:hypothetical Tubulin-tyrosine ligase containing protein, full insert sequence.
 DE library, clone:4930524K07 product:hypothetical Tubulin-tyrosine ligase containing protein, full insert sequence.
 OS Mus musculus (Mouse).
 GN Name=1700019P01Rik;
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Murioidea; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Testis;
 RC MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning."
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staib F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann C., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wuzhew-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kotsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RA The FANTOM Consortium,

the RIKEN Genome Exploration Research Group Phase I & II Team;
 "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs";
 RL Nature 420:563-573(2002).
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes."
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kitsumai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer."
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akaira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toy T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK029745; BAC36595.1; -; mRNA.
 DR Ensembl; ENSMUSG0000002388; Mus musculus.
 DR MGI; MGI:1922902; 1700019P01Rik.
 DR GO; GO:0016874; F:ligase activity; IEA.
 DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
 DR GO; GO:0006464; P:protein modification; IEA.
 DR InterPro; IPR004344; Tub_tyr_ligase.
 DR Pfam; PF03133; TTL; 1.
 KW Hypothetical protein; Ligase.
 SQ SEQUENCE 518 AA; 59319 MW; 540C045659FAC0D6 CRC64;
 Query Match 60.5%; Score 1726.5; DB 2; Length 518;
 Beat Local Similarity 71.4%; Pred. No. 2.1e-111;
 Matches 325; Conservative 57; Mismatches 64; Indels 9; Gaps 4;
 QY 1 MASSTLKWVSHQS-CSRSSRSKPRDQREAEAGSSDLSSRODAENAEALRGLPGQLVDIA 59
 DB 59 VAASILKVVVLHQNYC---SKYKGSKEEAKNSDPSPKDPENPDLKPLSLSGQVDDTA 115
 QY 60 CKVQAVLQGLQHEHEDIDTSADAVEDLTAEAEWEDLTQQYVSLVHGDAFTSNRSNYFSQOQA 119
 DB 116 CKVQAVLQGLQHEHEDIDVSEASTALSEEWNLDLTQQYVLLVHGNAFTDSKSYFAQOQA 175
 QY 120 LLNRITSVNPQTDIGLNNIWIIPAKSRGRDVCMDRVEEILEALAAADHPLSRDNKW 179
 DB 176 LLSKISSVNPQTEIDGIRNIWIIPAKSRGRDVCMDRVENILSLVAADSTQTKDNKW 235
 QY 180 VQKIETPLLICDTKFDIRQWFLVTDWNPPLTWFKESYLFSTQRFSLDKLDSAIHLCN 239
 DB 236 VQKIETPLLIYDTKFDIRQWFLVTDWNPPLTWFKESYLFSTQRFSLDKLDSAIHLCN 295

```
QY 240 NAVQKYLKNDVGRSPLLPAHNMWTSRFQYLQGRGAVMGSVIYPSMKKAIAHAMKVA 299
Db 296 NSHQRLKNDKERSPLPCHNMWTSRFQYLQGRGVTGWSIIYPSMKRAVINMRVA 355
QY 300 QDHVEPRKNSFELYGADFVLGRDFRPLWLEINSSPTMHPSTPTVTAQCAQVQEDTIKVA 359
Db 356 QDHVEARKNSFELYGADFVLGRDFKPLWLEINSSPTMHPSTPTVTAQCAQVQEDTIKVV 415
QY 360 -----DRSCDIGNFELLWRQVPEPPPSGSDLCVAGSVRRARQVLPVNCNLKASALLD 415
Db 416 DRKLDNRCDTGNFELLWRQVPEPPPSGSDLCVAGSVRRARQVLPVNCNLKASALLD 475
QY 416 AQLPKARGPSAMPDPAQGPSPALORDLGLKEKG 450
Db 476 APP-KVRSARALMETVIRPRTTVRQDWKREAKG 509

RESULT 4
Q8BV51_MOUSE
ID QB8V51_MOUSE PRELIMINARY; PRT; 704 AA.
AC QB8V51;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
DE library, clone:A630053H17 product:HOTT1 PROTEIN homolog.
GN Names=4833441J24Rik;
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Akaiwa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
```

```
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,
RA Sumi N., Iehi Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akaiura S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RC EMBL; AK080321; BC037878.1; -; mRNA.
DR Ensembl; ENSMUSG00000030276; Mus musculus.
DR MGI; MGI:2141418; 4833441J24Rik
DR GO; GO:0004835; P:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_ligase.
DR Pfam; PF03133; TTL; 1.
DR SQUENSE 704 AA; 79080 MW; 3FAD889C1DB5CF7D CRC64;
```

```
Query Match 36.3%; Score 1052; DB 2; Length 704;
Best Local Similarity 40.1%; Pred. No. 2.4e-64;
Matches 226; Conservative 85; Mismatches 177; Indels 76; Gaps 8;

QY 2 ASSILKVVVSHQSCRSRSRSPKPRDQREAGSSDLSSRQDAENAEAKLGLPGQLVDIACK 61
Db 119 ARNVKLVLKLEKSSQISIQARE--EEAPEDTQPKQ-----EKKLVTVSSDFVDEALS 171
QY 62 VCAQYLQGLQHEHEDIDTSADAVEDELTAEAWEDLTQQYVSLVHGDAFINSRNYFSQOALL 121
Db 172 ACQEHLSIIAHKDIKDPNSPLVSPDDWSQFLQRYQIVHEGAELRYLEVQVQCEDIL 231
QY 122 NRITSVNPTQDIDGLRNIIWIKPAKGRGDIIVCMRDVEILLELAADHPLSRDNKVVQ 181
Db 232 QQLQNVVPLQDMEDGRNIWIKPKAKSRGRGIMCMNRLDMLKLVDVNCNMLMKDGRKIVQ 291
QY 182 KYLETPLLICDTKFDIFQWFLVTDNPLTIWFKYESYLFSTQRFSLDKLDSALHLCNNA 241
Db 292 KYLERPLLIIFGTKFDLQWFLVTDNPLTVWFYRDSYIRFSTQFSLKNDLSVHLNNS 351
QY 242 VQKYLKNDVGRSPLLPAHNMWTSRFQYLQGRGAVMGSVIYPSMKKAIAHAMKVAQD 301
Db 352 IQRHLEASCHRHPLPDPNNWSSQRFQAHLEQVDPAKAWSSVIVPGMKAAVIAHALQTSQD 411
QY 302 HVEPRKNSFELYGADFVLGRDFRPLWLEINSSPTMHPSTPTVTAQCAQVQEDTIKVA 359
Db 412 NVQCRKASFELYGADFVFGEDFQPLWLEINASPTMAPSTAVTARLCAGVQADTLRVVIDR 471
```

```
QY 360 --DRSCDIGNFELLWRQPVVPPFSGSDLCVAGSVRRARQVLPVCMNKASASLLDAQ 417
Db 472 RLDSCDGTGAFELIYKQPAVEVPQYVGIRLLVEGTTIKK-----510
QY 418 PLKARGPSAMPDPAQGPSPALQORDLGLKEEKLPLALLAPLRGAESGGAOQPTRTKAA 477
Db 511 -----PVPVGHRTGVRS-----LPHLLTQGGESKDSGSPTH-RGA 548
QY 478 GKVELPACPCRHHVDSQAQNTGVVPAQF-----AKSWDPNQLNAHPLEPVL-- 523
Db 549 SRKNARAESLEHTEKPEPAASVSGKGKAPHPFSLHSKAWLPSPVRHPRQGRVLRQ 608
QY 524 -----GLKTAEGALRPPPGCK 539
Db 609 HDQLVGSKALSTTGKALMTLPTAK 632

RESULT 5
Q9H876 HUMAN
ID Q9H876_HUMAN PRELIMINARY; PRT; 744 AA.
AC Q9H876;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-WAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ13898.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Thyroid gland;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA Nagahori K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hirao S., Chiba Y.,
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,
RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA Nomura Y., Togiani S., Komai F., Hara R., Takeuchi K., Arita M.,
RA Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T., Nakagawa K.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45 (2004).
DR EMBL; AK03960; BAB14741.1; -; mRNA.
DR Ensembl; ENSG00000156983; Homo sapiens.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_lygase.
DR Pfam; PF03133; TTL; 1.
KW Ligase.
```

```
FT NON_TER 744 744
SQ SEQUENCE 744 AA; 84683 MW; DF661753E4FAFFODF CRC64;

Query Match 36.6%; Score 1043.5; DB 2; Length 744;
Best Local Similarity 40.5%; Pred. No. 1e-63;
Matches 234; Conservative 85; Mismatches 192; Indels 67; Gaps 12;

QY 2 ASSILKWVYSHQSCSRSSKPRDQREAGSDSLSRQDAENAEAKLGLGLGOLVADIACK 61
Db 151 ARNVKLKLVKSE-----WKSYPFQAVEEASGDQPKQEKPNVL---VSPEFVDEALC 201
QY 62 VQAYILQLEHEDIDTSADAVEDLTAEWEDLTQYIYSLVHGDAFTISNRYFSQOALL 121
Db 202 ACEEYLSNLAHMDIDKLEAPLYLTPEGWSLFLQRYIYVYVHEGAELHLDLTQVQRCEDIL 261
QY 122 NRITSVNPOTDIDGLRNIIKPAKSRGRDITVCMRVEEILELAAAHPLSRDNKVVVQ 181
Db 262 QQLQAVVFOIDMEGRNIIWIVPGAKSRGRGIMCDHLEMLKLVNGNPVVMKDKGVVQ 321
QY 182 KYIETPLLCIDTQDIQWFLVTDNPLTIWFKESYLFSTQRFSLDKLDSAIHLNNA 241
Db 322 KYIERPLLIFFTGLRQWFLVTDNPLTVWFYRDSYIRFSTQPFSLKNLDSNVHLNNS 381
QY 242 VQYILKNDVGRSPLLPANHWITSTRPQYLRQGRGAVGWSVITYPSMKKAIHAAMKVAQD 301
Db 382 IQKHLNSCHRRPLLPDDNMWSSQRFQHLQEMGAPNAWSTIIVPGMKDAVIHALQTSQD 441
QY 302 HVEPRKNSPELVGADPVLGRDFRPMLEINSPTWHPSTPVTQAQLCAQVQEDTIKAV-- 359
Db 442 TVQCRASFELYGADPVFGEDEFQPMLEINASPTWAPSTAVTARLCAGVQADLRVIDR 501
QY 360 --DRSCDIGNFELLWRQPVVPPFSGSDLCVAGSVRRARQVLPVCMNKASASLLDAQ 417
Db 502 MLDNRCDTGAFELIYKQPAVEVPQYVGIRLLVEGTTIKK-----PMACHRRMGVPAV 555
QY 418 PLKARGPS-----AMPDPAQGPSPALQORDLGL-----KEEK 450
Db 556 PLLTQRGSGEARHFPFLHTKQA-LPSPHVLRHQGVLRQHSKLVGTKALSTTGKALRT 614
QY 451 LPLA-----LLAPLRGAESGGAOQPTRKA-----AGKVELPACPCRHHVDSQ--AP-NTG 498
Db 615 LPTAKVFISLPLNDFKVPASITLKPRKPAALCLRGQPQLEVPCCCLCKSEQFLAPVGRS 674
QY 499 VVPAQPAKSWDPNQNLNAHPLEPVLRLGLTKAEGALRPPP 536
Db 675 RPKANSRPDCDKPRAEACPMKEL-----SPLKPLP 704

RESULT 6
Q9NDN8 HUMAN
ID Q9NDN8_HUMAN PRELIMINARY; PRT; 434 AA.
AC Q9NDN8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein DKFPZ586B0320.
GN Name=DKFPZ586B0320;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Uterus;
RA Wambutt R., Heubner D., Mewes H.W., Cassenhuber J., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=PCR rescued clones;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
```

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J.C., Hsieh F.,
 RA Dlatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Capleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hsieh S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield J.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RC NUCLEOTIDE SEQUENCE.
 RC TISSUE=PCR rescued clones;
 RG NIH MGC Project;
 RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL333939; CAB38794.1; -; mRNA.
 DR EMBL; BC098298; AA98298.1; -; mRNA.
 DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
 DR GO; GO:0006464; P:protein modification; IEA.
 DR InterPro; IPR004344; Tub_tyr_ligase.
 DR Pfam; PF01133; TTL; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 434 AA; 94933 MW; 9E79E6CA08651C1A1 CRC64;

Query Match 34.4%; Score 981.5; DB 2; Length 434;
 Best Local Similarity 46.7%; Pred. No. 9.9e-60;
 Matches 194; Conservative 64; Mismatches 106; Indels 51; Gaps 5;

QY 74 DIDSADAVEDLTEAEWEDLTQYYSLVHGDAFISNSRNYFSQOALLNRITSVNPOTDI 133
 DB 2 DIDKDLAPLYLTPEGWSLFLQRYQVHGAEALRLDVTQVQRCEDILQQLQAVVPQIDM 61
 QY 134 DGLRNIWIIPKAASRGSRDIVCMRDVEEILELAADHPLSRDNKVVQVQKIETPLLCIDT 193
 DB 62 EGRNIWIVPGAKSRGSGIMCMHLEMLKLVNGPNVVMKDGKVVQVQKIETPLLCIDT 121
 QY 194 KFDIOWFLVTDWNPITWIFKESYLFSTQSFSLDKLDSAIHLCNNAVOKYLVNDVGRS 253
 DB 122 KFDLRQWFLVTDWNPITWIFKESYLFSTQSFSLDKLDSAIHLCNNAVOKYLVNDVGRS 181
 QY 254 PLLPAHNMWTSRFOEVLQRCGAVGWSVIYPSMKKAIHAAMKVAADHVEPRKNSPELY 313
 DB 182 PLLPPDNMSSQRFQALHENGAPNAWSTIIVPGMDAVIHALQTSQDTVQCRKASPELY 241
 QY 314 GADPVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAV----DRSCDIGNFE 369
 DB 242 GADPVGEDPQPLWIEINASSPTMAPSTAVTARLCAGVQADTLRVVIDRLDRNCDTGAFE 301
 QY 370 LLMRQPVVEPPFPFSGDLVAGSVRRARRQVLPVNCNKASASLLDAPQLKARGPSAMD 429
 DB 302 LIYQPAVEPQVYGIIRLLVEGFTIKK-----PMAM-- 332
 QY 430 PAQPPSPALQDGLGKBEKGLPLALLAPLRGAESGGAQPTTKAAGKVELPA 484
 DB 333 -----CHRRMGVR-----PAVPLLTQRGSGEARHHPFLSLTKA-----QLPS 369

RESULT 7
 Q6AWA3_HUMAN
 ID Q6AWA3 HUMAN PRELIMINARY; PRT; 352 AA.
 AC Q6AWA3;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Hypothetical protein DKFZp686D076.

GN Name=DKFZp686D076;
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RC NUCLEOTIDE SEQUENCE.
 RC TISSUE=Cervix;
 RG The German cDNA Consortium;
 RA Ansoorge W., Krieger S., Regier T., Rittmüller C., Schwager B.,
 RA Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
 RA Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
 RL EMBL; BX648175; CAH10554.1; -; mRNA.
 DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
 DR GO; GO:0006464; P:protein modification; IEA.
 DR InterPro; IPR004344; Tub_tyr_ligase.
 DR Pfam; PF01133; TTL; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 352 AA; 40257 MW; 49FD8E9CAE1CB20D CRC64;

Query Match 33.5%; Score 957.5; DB 2; Length 352;
 Best Local Similarity 54.2%; Pred. No. 3.5e-58;
 Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;

QY 74 DIDSADAVEDLTEAEWEDLTQYYSLVHGDAFISNSRNYFSQOALLNRITSVNPOTDI 133
 DB 2 DIDKDLAPLYLTPEGWSLFLQRYQVHGAEALRLDVTQVQRCEDILQQLQAVVPQIDM 61
 QY 134 DGLRNIWIIPKAASRGSRDIVCMRDVEEILELAADHPLSRDNKVVQVQKIETPLLCIDT 193
 DB 62 EGRNIWIVPGAKSRGSGIMCMHLEMLKLVNGPNVVMKDGKVVQVQKIETPLLCIDT 121
 QY 194 KFDIOWFLVTDWNPITWIFKESYLFSTQSFSLDKLDSAIHLCNNAVOKYLVNDVGRS 253
 DB 122 KFDLRQWFLVTDWNPITWIFKESYLFSTQSFSLDKLDSAIHLCNNAVOKYLVNDVGRS 181
 QY 254 PLLPAHNMWTSRFOEVLQRCGAVGWSVIYPSMKKAIHAAMKVAADHVEPRKNSPELY 313
 DB 182 PLLPPDNMSSQRFQALHENGAPNAWSTIIVPGMDAVIHALQTSQDTVQCRKASPELY 241
 QY 314 GADPVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAV----DRSCDIGNFE 369
 DB 242 GADPVGEDPQPLWIEINASSPTMAPSTAVTARLCAGVQADTLRVVIDRLDRNCDTGAFE 301
 QY 370 LLMRQPVVEPPFPFSGDLVAGSVRRARRQVLPVNCNKASASLLDAPQLKARGPSAMD 429
 DB 302 LIYQPAVEPQVYGIIRLLVEGFTIKK-----PMAM-- 332

RESULT 8
 TTL3_HUMAN
 ID TTL3 HUMAN STANDARD; PRT; 352 AA.
 AC Q9Y4R7; Q9UI99;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)
 DE Tubulin tyrosine ligase-like protein 3 (HOTT1).
 GN Name=TTL3; ORFNames=PRO207;
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RC NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RC TISSUE=Testis;
 RG The German cDNA consortium;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 61-352.
 RC TISSUE=Fetal liver;
 RA Yu Y., Zhang C., Luo L., Ouyang S., Zhang S., Li W., Wu J., Zhou S.,


```

OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RP Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud C., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segures B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Bremond C., Skalli Z., Cattolico L., Poulin J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RN Nature 431:946-957 (2004).
[2]
RN NUCLEOTIDE SEQUENCE.
RP Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR ENBL; CAAG0104979; CAG06724.1; -; Genomic DNA.
SQ SEQUENCE 534 AA; 60130 MW; 62B5BBB696B662B CRC64;

Query Match 33.0%; Score 941.5; DB 2; Length 534;
Best Local Similarity 49.2%; Pred. No. 7.9e-57;
Matches 192; Conservative 59; Mismatches 118; Indels 21; Gaps 6;

QY 3 SSILKVVVSHQSCRSRSRKPQREAGSDLSRODAENAEAKLGLPGQLVDITACKV 62
DB 146 TSLLOHVVEGSSHRA-----DAKEE---QLHLPEDL--SKHLFPFNAGVIATLRV 193
QY 63 CQAYLGQLEHEDITSADAVEDLTEAWEDLTQOYSLVHGDAFISNRNYSQOQALLN 122
DB 194 CQHYLGVLGHEHROIDVT---METGEVSWGAFLODYKRVHVGASVGGGVFERCQAMLH 250
QY 123 RITSVNPQTDIDGLRNIIWIKPAKSRGRDIVCMDRVVEEILEAAADHPLSRKNKVVQK 182
DB 251 KLQEVCEQLGTDGLNIIWIKPGMSRGRGIVCKRDLDEILAVDNDKSLNKEKKVVQK 310
QY 183 YIETPLLCDTKFDIRQWFLVDNPLTIWFKYESYLRFSTQRFSLDKLDSAIHLGNVAV 242
DB 311 YLERPLLAVHGTGKFDLRQWFLVDNPLTIWFWFYRECYLRFSTQFPSTKILDSIHLNCSI 370
QY 243 QKYLKNDVGRSPLLPAHNMWTSRFBQYLQQRGAVGWSVIVPSMKKATAHAMKVAODH 302
DB 371 QKHFEPAARNRHPAVPGDNMWSQSFRAFLHQQGRSAEWVTVFGMQAKIIRALQTAQEQ 430
QY 303 VEPKNSFELYGADFVLGRDPRPWLIEINSPTHPSTPTVTAQICAOVQEDTIKVAVDNR 362
DB 431 VEPKNSFELYGADFVLMGRDLRPLMLLEINICPTMAQSSVVTALHCLPAVQIDTLRVLLDR 490
QY 363 CD-----IGNFELLWRQPVPEP--PPFSGSD 386
DB 491 SDPSAHTGGFOLICKRAGGSGGCPLFSGSE 520

RESULT 11
Q922T0_MOUSE
ID Q922T0_MOUSE PRELIMINARY; PRT; 266 AA.
AC Q922T0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

```

```

DE 4833441J24Rik protein.
GN Name=4833441J24Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Malek J.A., Gunaratne P.H.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Richards S., Worley K.C., Hale S., McKernan K.J., Green E.D., Dickson M.C.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RG NIH MGC Project;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
EMBL: BC006830; AAH06830.1; -; mRNA.
DR Ensembl; ENSMUSG00000030276; Mus musculus.
DR MGI; MGI:2141418; 4833441J24Rik.
DR GO; GO:0004835; F:ubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_ligase.
DR Pfam; PF03133; TTL; 1.
DR PFAM; PF03133; TTL; 1.
SQ SEQUENCE 266 AA; 30506 MW; F8E8FB52FA8B8E98 CRC64;

Query Match 29.2%; Score 834; DB 2; Length 266;
Best Local Similarity 60.6%; Pred. No. 9.1e-50;
Matches 149; Conservative 44; Mismatches 49; Indels 4; Gaps 1;

QY 133 IDGLRNIIWIKPAKSRGRDIVCMDRVVEEILEAAADHPLSRKNKVVQKIETPLLCD 192
DB 1 MEGDQNTIWKPKAKSRGRGIMCNRLDEMLKLVDCNPLMKOGKVIQKYEPLLIIFG 60
QY 193 TKEDIRQWFLVDNPLTIWFKYESYLRFSTQRFSLDKLDSAIHLGNVAVQKYLKNDVGR 252
DB 61 TKEDLRQWFLVDNPLTIWFWFYRDSYRFSYRFSYRFSYRFSYRFSYRFSYRFSYRFSYR 120
QY 253 SPILLPAHNMWTSRFBQYLQQRGAVGWSVIVPSMKKATAHAMKVAODHVEPRKNSFEL 312
DB 121 HPMLPPDNMWSQSFRAFLHQQGRSAEWVTVFGMQAKIIRALQTAQEQ 180
QY 313 YGADPVLGRDPRPWLIEINSPTHPSTPTVTAQICAOVQEDTIKVAV----DRSCDIGNF 368
DB 181 YGADFVFGEDFQPLWLEINASPTMAPSTAVTARLCAGVQADTLKRVDRRLDRSCDTGAF 240
QY 369 ELLWRQ 374
DB 241 ELIYKQ 246

RESULT 12
Q58CT2_BOVIN
ID Q58CT2_BOVIN PRELIMINARY; PRT; 261 AA.

```


RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Banton R.C., Rogers J.-H.C., Blazet R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Franckoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
[2]
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=22426055; PubMed=12537568;
RX Celinker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
[3]
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=22426070; PubMed=12537573;
RX Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celinker S.E.;
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
RT a genomic perspective.";
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
[4]
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=22426069; PubMed=12537572;
RX Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celinker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).

RN NUCLEOTIDE SEQUENCE.
RP Berkeley *Drosophila* Genome Project;
RA Celinker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirskas R., Smith E.,
RA Yu C., Rubin G.;
RT "Drosophila melanogaster release 4 sequence.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[6]
RN NUCLEOTIDE SEQUENCE.
RP FlyBase;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL: AS003614; AAF52432.1; -; Genomic DNA.
DR Ensembl: CG11323; *Drosophila melanogaster*.
DR FlyBase: Fgn0031854; CG11323.
DR GO: GO:0004835; *F*tubulin-tyrosine ligase activity; IEA.
DR GO: GO:0006464; *P*rotein modification; IEA.
DR InterPro: IPR004344; *Tub* tyr_lysase.
DR Pfam: PF03133; *TTL*; 1.
SQ SEQUENCE 992 AA; 112255 MW; 0FC34674C15BDAAD CRC64;
Query Match 25.3%; Score 722.5; DB 2; Length 992;
Best Local Similarity 33.3%; Pred. No. 3e-41;
Matches 175; Conservative 86; Mismatches 198; Indels 67; Gaps 12;
QY 51 LPQQLVDIAKVCQAVLGQLEHEDIDTSADAVED---LTEAEWEDLTQQYSLVHGDAFI 107
DB 326 IPYSAIDFAYKRLVEYIDSCQHNID-----FEDPKIWEHDWDAFLFQHQQLVNEDGRI 380
QY 108 SNS-----RNYFSQCQALLNRNITSVNPQTDIDGLRNIIKPAKSGRGRDIVCMRVVEI 162
DB 381 QHGGQRLSPWVKSLSLVDKMKVHPQYSLDGYQNNWIKVPANKKCRGRIILMDNLKKI 440
QY 163 LELAAADHPLSRDNKWKVQKIETPLLCIDTFDIQWFLVTDWNLPLTWFKESYLRF 222
DB 441 --LGVVNLSTASKSRVYVQKIERPLILFQTKEDIRQWFLTNTQPLVVMFYRESYLRF 498
QY 223 TQRFSLDKLDSAIHLCNNAVQKYLKNDVGRSPLPAHNNWMTSTRFOYLQROGRGAVWGS 282
DB 499 SQEYSLSNHHESVHLTNYAIQKKYNG-KRDKLPSNMMWDCYSFOAYLKQIKYNNWLE 557
QY 283 VIYPSMKKATAHAMVAQDHVEPRKNSFELYGADFVLGRDFRPLWIEINSSPTMHPSTPV 342
DB 558 RIFPGMKALVGMCLASQENMDRPRPTFELFGADFMCENFYFWLIEINSSPDLGATTSV 617
QY 343 TAQLCAQVQBDTIKVAVDKSCD-----IGNFELLWRQPVVEPPFPFSGSDLCVAGSVV--- 394
DB 618 TARMCPQCLDDVVKVWIDRRTPDKAELGNFELAYRQVVPPTPAYMGLNLFVKQKVLQKA 677
QY 395 -----RRARQVLPVCNLKASASLDAQPLKARGPSAMP----- 428
DB 678 NHGGHGHVYQQQKERSLATSSVYRQRSAIHPATSIIRHAMPTFNATEYMEKVMV 737
QY 429 DPAQGPSPALQDLGLKEEKLGLFALLAPLRGAESGGAQPTRTKAAGKVELPACPCR 488
DB 738 EPLSSRS-SLCQLPKQSPSAAPALTATP-----SGATSSYLKQAGRSITQLLSAT 789
QY 489 H-----VDSQAPNTGVPAQPAKSWDPNQLNAHPLEPVLRLGLK 526
DB 790 HKRNTGGLSGEQVQSTALP-PKRQRSCGPLRSLSTNPVESTKKFK 834
RESULT 15
Q7PMD3 ANOGA
ID Q7PMD3 ANOGA PRELIMINARY; PRT; 501 AA.
AC Q7PMD3;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ENSANGP0000022334 (Fragment).
GN ORFNames=ENSANG00000019845;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Search completed: April 4, 2006, 12:45:56
Job time : 82.9545 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:45:31 ; Search time 20.1115 Seconds
(without alignments)
2588.231 Million cell updates/sec

Title: US-10-635-977-2
Perfect score: 2854
Sequence: 1 MASSILKVVVSHQSCSRSR.....LRGLKTAEGALRPPPGKGS 541

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR.80.*
2: PIR1.*
3: PIR2.*
4: PIR3.*
5: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	956.5	33.5	352	2 T12515	hypothetical prote
2	259	9.1	379	2 A45443	tubulin-tyrosine 1
3	252.5	8.8	640	2 E88575	protein ZK1128.6 [
4	252.5	8.8	680	2 T27699	hypothetical prote
5	209.5	7.3	1198	2 T20262	hypothetical prote
6	209.5	7.3	1203	2 C89217	protein C55A6.2 [i
7	166	5.8	662	2 T20343	hypothetical prote
8	161	5.6	403	2 T37571	tubulin-tyrosine 1
9	155.5	5.4	883	2 A96805	hypothetical prote
10	144.5	5.1	753	2 S48261	hypothetical prote
11	130	4.6	2472	2 E83594	still frameshift p
12	124	4.3	4660	2 T42737	gp330 protein prec
13	120	4.2	882	2 A70507	probable reductase
14	119.5	4.2	3436	2 S55659	tegument protein 6
15	115.5	4.0	1006	2 T41104	probable transcrip
16	114.5	4.0	1336	2 S25716	Ras guanine nucleo
17	112.5	3.9	696	2 T45394	hypothetical prote
18	112.5	3.9	1575	2 S68448	synaptotagmin, 170K
19	111	3.9	930	2 A25923	progesterone recep
20	109	3.8	906	2 A43817	transforming prote
21	108.5	3.8	741	2 I51657	suppressor of yeas
22	108.5	3.8	1048	2 T31425	C-terminal domain
23	108.5	3.8	1650	2 S53457	dominant autoantig
24	108.5	3.8	2176	2 T39188	probable U5 snRNP
25	108	3.8	940	2 JE0291	FB19 protein - hum
26	108	3.8	1245	2 G84897	hypothetical prote
27	108	3.8	3149	1 Q9BE8	BPLF1 protein - hu
28	107.5	3.8	420	2 T46910	hypothetical prote
29	107.5	3.8	896	2 B43817	transforming prote

30 107 3.7 755 2 S32103
31 107 3.7 1234 2 T00363
32 106.5 3.7 538 2 S57459
33 106.5 3.7 2870 2 H96974
34 105.5 3.7 517 2 E89530
35 105.5 3.7 4957 2 T03455
36 105.5 3.7 5262 2 T03454
37 104 3.6 836 2 T32298
38 103.5 3.6 1290 2 T00018
39 103.5 3.6 2774 2 A43359
40 103 3.6 670 2 D86176
41 103 3.6 816 2 T15049
42 102.5 3.6 775 2 S63626
43 102.5 3.6 906 2 JCS963
44 102 3.6 454 2 A84162
45 102 3.6 454 2 T08280

ALIGNMENTS

RESULT 1

T12515
hypothetical protein DKFZp434B103.1 - human
C:Species: Homo sapiens (man)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 09-Jul-2004
C:Accession: T12515
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z17524
A:Accession: T12515
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-352 <WAM>
A:Cross-references: UNIPROT:Q9Y4R7; UNIPARC:UPI00001377A6; EMBL:AL096725
A:Experimental source: adult testis; clone DKFZp434B103
C:Genetics:
A:Note: DKFZp434B103.1

Query Match 33.5%; Score 956.5; DB 2; Length 352;
Best Local Similarity 54.2%; Pred. No. 78-64; Indels 5; Gaps 2;
Matches 179; Conservative 55; Mismatches 91;
QY 74 DIDSADAVDELTAEWEDLTQQYSLVHGDAFISNRNYFQCOALLNRITSVNPQTDI 133
Db 2 DIDKLEAPLYLTPEGWSLFLQRYQVHVHGAELSHLDTQVQRCEDILQLOAVVPQIDM 61
QY 134 DGLRNIIWKPAKSGRGDIIVCMDRVVEILEAAADHPLSRDNKVVQKYIETPLICDT 193
Db 62 EGDRIWIVKPKAKSGRGIMCMCHLEMLKLVNGNPNVVMKDKGVQKYIERPLIFGT 121
QY 194 KFDIQWELVTDWNPFLTITWEYKESYLRFSTQSFSDKLSATHLCNNAVQKYLKNDVGRS 253
Db 122 KFDLQWELVTDWNPFLTITWYRDSYIRFSTQSFSLKLNLDNSVHLNNSIQKHLNSCHRH 181
QY 254 PLLPANNMTSTRFQYELQQRGAVGWSVTPYSMKKAIAHAMKVAQDHVEPRKQSFELY 313
Db 182 PLLPDDNMSSQRFQAHQENGAPNAWSTIIVPGMKDAVIHALQTSQDTVQCRKASFELY 241
QY 314 GADFVLGRDFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKAV----DRSCDIGNPE 369
Db 242 GADFVFGEDFQFWLIEINASTMAPSTAVTARLCAGVQADTLRVVIDRMLDRNCDTGAFE 301
QY 370 LLWRQPVVEPPFPFGSDLCVAGV-SVRRAR 398
Db 302 LIYQPVTTSPASTPRPSCLLPMSYDTRAR 331

RESULT 2

A45443
tubulin-tyrosine ligase (EC 6.3.2.25) [validated] - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C;Accession: A45443
R;Brsfeld, K.; Wehland, J.; Plessemann, U.; Dodemont, H.; Gerke, V.; Weber, K.
J. Cell Biol. 120, 725-732, 1993
A;Title: Characterization of the tubulin-tyrosine ligase.
A;Reference number: A45443; MUID:93147125; PMID:8093886
A;Accession: A45443
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-379 <ERS>
A;Cross-references: UNIPROT:P38160; UNIPARC:UPI00001377A9; GB:X68453; GB:S54050; NID:921
A;Experimental source: brain
A;Note: sequence extracted from NCBI backbone (NCBI:123858, NCBI:123859)
C;Keywords: ligase

Query Match 9.1%; Score 259; DB 2; Length 379;
Best Local Similarity 26.8%; Pred. No. 1.2e-11;
Matches 98; Conservative 57; Mismatches 128; Indels 82; Gaps 20;
Qy 52 PG---QLVDI---ACKVC-QAYLQLEHEDITDSADAVEDLTEAEWEDLTQOYYSL----- 100
Db 56 PGLMQLVNYVRGADKLCRKASLVKLT---IKTSPLEAECT---WFPESYVIYPTNLKTP 108
Qy 101 -----VHCDAFISNRVYFSCQALLNRITSNPQTDIDGLRNWIWIKPAKSR 149
Db 109 VAPAQNGIHPHSSR--TDREFF-----LTSYNKKKE--DGEQNVIAKSSAGAK 156
Qy 150 GRDITVCMRDRVEEILEAAADHPLSRDNK---WVQKYIETPLLI--CDTKFDIRQWFLVT 204
Db 157 GEGLLISSEATELDDFI-----DNQGVHVQKYQLERPLLEPGHGRKFDIRSWVLVD 208
Qy 205 DMPPLTTFYKESVLRSTORFSLDKL-DSAIHLCNNAVOKYLKNDVGRSFLPLPAHNMWT 263
Db 209 --HOYNIYLYREGVLRRTASEPYHTDNFDQDKTCHLTNHCIOKEYSKNYGK---YEEGNEFM 263
Qy 264 STRQEVLRQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPR-----KNSFELYGAD 316
Db 264 FEEFNQVLT-----SALNITLESILQIKHIRSCILLSVEPALSTRHLPYQSFQLRGFD 318
Qy 317 FVLGRDRFRLMIEINSSPTMHPSPFVTAQLCAQVQEDTIKVAVDRSCDIGNFELLWRQPV 376
Db 319 FWVDEDLKWVLEVNGAPAC--AQKLYAELC---QGIVDIAIASVFPDPDAE----QQQ 368
Qy 377 VEPPP 381
Db 369 QQPPP 373

RESULT 3
E88575
protein ZK1128.6 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C;Accession: E88575
R;anonymous, the C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: E88575
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-640 <STO>
A;Cross-references: UNIPROT:Q09647; UNIPARC:UPI000017A077; GB:chr_III; PIDN:CAA87425.1;
C;Genetics:
A;Gene: ZK1128.6
A;Map position: 3

Query Match 8.8%; Score 252.5; DB 2; Length 640;
Best Local Similarity 25.6%; Pred. No. 7.6e-11;
Matches 81; Conservative 56; Mismatches 118; Indels 61; Gaps 12;
Qy 132 DIDGLRNWIWIKPAKSRGRDIVCMRDRVEEILEAAADHPLSRDNKWWVQKYIETPLLIC 191

Db 283 ETDASRHHV--IVKPPASARGTGISVTRKPKDPPTATL-----VAQHYIERPLTIN 331
Qy 192 DTKFDIRQWFLVTDWNPPLTIWFYKESYLRSTORFSLDK---LDSAIHLCNNAVOKYLKN 248
Db 332 RAKFDLRLYAYVPTFEPLRVYIDQGLVRFAFVPSVSHSVSTISNKYMHLTNSINKLAEA 391
Qy 249 D-VGRSPL--LPAHNMWTSTRF-----QEYLQQRGAVGWSVIYPSMKKAIAHAMKVA 299
Db 392 DGVANKFPVKWTLHLHLW--EHFDEMVGVDREKIQRE-----IEEVIKAFISTEKPI 440
Qy 300 QDH---VEPRKNSFELYGADFVLGRDPRFWPLIEINSSPTMHPSTPTVTAQLCAQVQEDTI 355
Db 441 REHMSRFLQEFICYELFGDIIIDEDYKFWLLEVNISPSLHSGTPLDVSVKAPLAKDVL 500
Qy 356 KVAVDRSCDIGNFELLWRQPVVEPPPF---SGSDLCVAGSVRRARRQVLP-----VCNLK 408
Db 501 NLA-----GVYVPPSFDKLSADADYSTPRNGRKNRQELIKEASVWAAYK 544
Qy 409 ASASILLDAQPLKARGP 424
Db 545 DQLGVIDNRIFKRLTP 560

RESULT 4
T27699
hypothetical protein ZK1128.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T27699
R;Barks, M.
submitted to the EMBL Data Library, January 1995
A;Reference number: Z20407
A;Accession: T27699
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-680 <MTL>
A;Cross-references: UNIPROT:Q09647; UNIPARC:UPI000017A076; EMBL:Z47357; PIDN:CAA87425.2
A;Experimental source: clone ZK1128
C;Genetics:
A;Gene: CESP:ZK1128.6
A;Map position: 3
A;Introns: 114/1; 208/3; 242/3; 334/3; 362/2; 405/2; 424/2; 550/2; 633/2

Query Match 8.8%; Score 252.5; DB 2; Length 680;
Best Local Similarity 25.6%; Pred. No. 8.3e-11;
Matches 81; Conservative 56; Mismatches 118; Indels 61; Gaps 12;

Qy 132 DIDGLRNWIWIKPAKSRGRDIVCMRDRVEEILEAAADHPLSRDNKWWVQKYIETPLLIC 191
Db 323 ETDASRHHV--IVKPPASARGTGISVTRKPKDPPTATL-----VAQHYIERPLTIN 371
Qy 192 DTKFDIRQWFLVTDWNPPLTIWFYKESYLRSTORFSLDK---LDSAIHLCNNAVOKYLKN 248
Db 372 RAKFDLRLYAYVPTFEPLRVYIDQGLVRFAFVPSVSHSVSTISNKYMHLTNSINKLAEA 431
Qy 249 D-VGRSPL--LPAHNMWTSTRF-----QEYLQQRGAVGWSVIYPSMKKAIAHAMKVA 299
Db 432 DGVANKFPVKWTLHLHLW--EHFDEMVGVDREKIQRE-----IEEVIKAFISTEKPI 480
Qy 300 QDH---VEPRKNSFELYGADFVLGRDPRFWPLIEINSSPTMHPSTPTVTAQLCAQVQEDTI 355
Db 481 REHMSRFLQEFICYELFGDIIIDEDYKFWLLEVNISPSLHSGTPLDVSVKAPLAKDVL 540
Qy 356 KVAVDRSCDIGNFELLWRQPVVEPPPF---SGSDLCVAGSVRRARRQVLP-----VCNLK 408
Db 541 NLA-----GVYVPPSFDKLSADADYSTPRNGRKNRQELIKEASVWAAYK 584

Qy 409 ASASILLDAQPLKARGP 424
Db 585 DQLGVIDNRIFKRLTP 600

A;Accession: T37571
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-403 <BAD>
A;Cross-references: UNIPROT:Q10438; UNIPARC:UPI000013A94B; EMBL:Z70721; PIDN:CAA94694.1;
A;Experimental source: strain 972h-; cosmid c12B10
C;Genetics:
A;Gene: SPDB:SPAC12B10.04
A;Map position: 1
A;Introns: 320/3; 348/3

Query Match 5.6%; Score 161; DB 2; Length 403;
Best Local Similarity 20.8%; Pred. No. 0.00027;
Matches 72; Conservative 65; Mismatches 127; Indels 86; Gaps 14;

QY 71 EHEIDITSADAV-----EDLTAEWEDLTQQYYSLVHGDAFISNS-----110
DB 49 QYEDID--FDEVYKPKTKLCCSVIRKALIRKEYLWRTVITYLAKHPDSILSKSVPEAY 106
QY 111 ---RTFSQCOALLNRTSVNPQTDIDGLRNI-----WIIPKAASGRGRDIVCMRVEEI 162
DB 107 SLELDYAEFLDDSLMEAYELRQLEENATKNISEKQWYILKPSMCDRAQGIRLFSTIEEL 166
QY 163 LEL-----AAADHPLSRDNK-----WVQKYIETPLLCDTKFDI 197
DB 167 QAIFSDFDDESESEAGLEEKGDIITVAFNKNIVISOIRNFLVQKYSKPLLLDHRKFI 226
QY 198 ROWFLVTWNPLTIWFKESYLFSTQFSLDKLDSAI---HLCNNAVQKYLKNDVGRSP 254
DB 227 RAVLAT--GALSVYLFENMLCLLARDKYKPTFPDPLFLSHLSNTCLO---GDNVQSS 281
QY 255 LLPAHNMWTSRFOEYLQRGGAVMGVSIVYPSMKKAIHAHAMKVAQD---HVEPRKNSF 310
DB 282 I---RDFWNTS-----IENK-----DDIFKSIILNIIGDVFEAAATTOGIHQPLENCF 326
QY 311 ELYGADFVLGRDRPRPMLIENSSFTWHPSTPTVAQLCAQVOEDTIKAVD 360
DB 327 EIRGVDFLVDCEQVYLLEVNS----YPDPKQTGKNLSNIENLFSAVVE 372

RESULT 9
A96805
hypothetical protein TSM16.14 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: A96805
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: A96805
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-883 <STO>
A;Cross-references: UNIPROT:Q9CAQ1; UNIPARC:UPI000000A09A3; GB:AB005173; NID:G6382502; PI
C;Genetics:
A;Gene: TSM16.14
A;Map position: 1

Query Match 5.4%; Score 155.5; DB 2; Length 883;
Best Local Similarity 25.3%; Pred. No. 0.0022;
Matches 55; Conservative 36; Mismatches 93; Indels 33; Gaps 7;

QY 134 DGLRNWIIPKAASGRGRDIVCMRVEIELELAADHPLSRDNKWWVQKYIETPLLCDT 193
DB 641 DQLNNWILKPNWARTIDTISDNLSAIIR-----MMETGPKICQKIETHPALPKGN 693

QY 194 KFDIROWFLVTDWNPDLTIWFKESYLFSTQFSLDKLDSAIHLCCNNAVQKYLKNDVGRS 253
DB 694 KFDLRYVVLVRSIDPLEIYLIIFWVRLSNPNYSLEKHSFFETHTFTVWNY-----GRK 748
QY 254 PLLPAHNMWTSRFOEYLQROG-----RGAVMGSV-----IYPSMK---KAIAHAMK 297
DB 749 ----LNHKPTAEFVREFEQEHNDYAFHFVNNTIQLSIVKWMIDIHEKVKQVIRAVFEAAA 804
QY 298 VAQDHVEPRKNSFELYGADFVLGRDRPRPMLIENSSP 334
DB 805 LAHPMQSPK-SRAMYGVDMVLDSSPEPKILEVTYCP 840

RESULT 10
S48261
hypothetical protein YBR094w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YBR0821
C;Species: Saccharomyces cerevisiae
C;Date: 01-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 09-Jul-2004
C;Accession: S48261; S45962; S41800; S44676
R;Mannhaupt, G.; Stucka, R.; Ehnlé, S.; Vetter, I.; Feldmann, H.
Yeast 10, 1363-1381, 1994
A;Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
A;Reference number: S48255; MUID:95208357; PMID:7900426
A;Accession: S48261
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-753 <MAN>
A;Cross-references: UNIPROT:P38254; UNIPARC:UPI000013A3DD; EMBL:X78993; NID:G476045; PI
R;Feldmann, H.; Mannhaupt, G.; Schwarzlose, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45927
A;Accession: S45962
A;Molecule type: DNA
A;Residues: 1-753 <PE2>
A;Cross-references: UNIPARC:UPI000013A3DD; EMBL:Z35963; NID:G536366; PIDN:CAA85047.1; PID
R;Dekker, P.J.T.; Hoekert, W.; van Oosterum, K.; Grivell, L.A.
submitted to the EMBL Data Library, December 1992
A;Reference number: S31313
A;Accession: S41800
A;Molecule type: DNA
A;Residues: 167-351, 'TPE', 356-449, 'A', 451-562, 'R', 564-753 <DEK>
A;Cross-references: UNIPARC:UPI0000168D37; EMBL:X69881; NID:G4086; PIDN:CAA49508.1; PID
C;Genetics:
A;Cross-references: SGD:S00000298
A;Map position: 2R

Query Match 5.1%; Score 144.5; DB 2; Length 753;
Best Local Similarity 21.8%; Pred. No. 0.012;
Matches 57; Conservative 44; Mismatches 105; Indels 56; Gaps 9;

QY 140 WIIPKAASGRGRDIVCMRVEEI-----LELAADHP-----171
DB 476 WIVKPSMSDKGGIRFKVTIEDIQAIFDSFDDESEAESGNDDDADDVNGVEFMNKNVN 535
QY 172 LSRDNKWWVQKYIETPLLIC---DTKFDIROWFLVTDWNPDLTIWFKESYLFSTQF-- 226
DB 536 ISQLRHFIIOEYLTNPILLASMDNRKFIRCY--VVCGRGLQVFFVDRMLALFAAKFPVP 593
QY 227 -----SLDKLDSAIHLCCNNAVQKYLKNDVGRSPILLPAHNMWTSRFO---EYLQROGR 277
DB 594 LDPYAVSVTDKLECHLTNTCLOQS-KKKDKDSVL-----EFDSEIEELPNERKS 642
QY 278 AVMGSVIYPSMKKAIHAHAMKVAQDHVEPRKNSFELYGADFVLGRDRPRPMLIENSSPMH 337
DB 643 NI-KEQIHSITNDVFLAAVNVNRLNFOPLNFAETYGVDFLDSNVEVKLEINAFPDFK 701
QY 338 PSTPVTVAQLCAQVOEDTIKAV 359
DB 702 QTGKDLKNLIDELFDDTVKYCV 723

RESULT 11

EB3594
still frameshift probable component of chemotactic signal transduction system PA0413 [in
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C:Accession: E83594
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: E83594
A:Molecule type: DNA
A:Residues: 1-2472 <STO>
A:Cross-references: UNIPROT:Q91696; UNIPARC:UPI00000C5045; GB:AE004478; GB:AE004091; NID
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0413

Query Match 4.6%; Score 130; DB 2; Length 2472;
Best Local Similarity 19.7%; Pred. No. 0.79;
Matches 122; Conservative 81; Mismatches 203; Indels 214; Gaps 29;

QY 71 EHHDIIT-----SADAVEDLTEAEWEDLTQQYYSLVHGDAFISNSRNYFSQCOALLN 122
DB DHEVTDAERARLAGPRDAMRSVVGALCBELVRKIDSL---DLFVRSDRGHPSELDAILA 354
QY 123 RITSVNPQTDIDG-----LRNIWIIPKAKSRGR--DIVCMDEVEIILSLAA----- 167
DB PLKQIADTLAVIGFGQPRKVIQDLQDVIIHALAQGRREPSDAIMVAGALLYVEATLAGM 414
QY 168 ---ADHPLSRDNKVVQKIYET-PLLICDTKFDIRO-----WFLVTDWNPILTIFWYKE- 216
DB AGPGDERNESESLPTDVAQIHQLVKEARNGLEQAKDAIIEFTASQWNHEHLARVPEL 474
QY 217 -SYLRFSTORFSLDKLSAIIHCNNAVQKYNVDVGRSPLLPAHNNWTS-----TRF 267
DB LQVRRGLLAMIPLERAATLTLCNRYIQBL---LARK-----AVPDWQSLDTLADAITSV 527
QY 268 QEYLORQGRGAVGVSIVYPSMKKATAHAKVQAQDHVEPRKNSFELYGADFVLRDRFPWL 327
DB EYLLERL-----SEDHASQSDLI---LDVAEDSLAN-----LGYTLKP-- 562
QY 328 IBINS-----SPTMHPSTPTVTAQLCAQ-----VOEDTIKVA 358
DB ---NSSAPAEPLGSGPAATESPAEPEPEAVVEVAETAEPADTAPAEAREADAPQLA 619
QY 359 VDRSCDIG-----NPELLWRQPV-----VEPPP-----F 382
DB SDNNWTLGKVADPAGEDPSLDLALDPLDLSAEVPPALPEVVEESGQPOSTSAPARSLDDF 679
QY 383 SSSDLGVAGSVRRARRQVLPVNCNLKASASILLDAQPLKARGPSAMPD----- 429
DB SLDEIDLSGLD-----LPA-----DAAAP--ASGPAALADWSLPEQWGLGGDL 719
QY 430 --PAQGPS-----PALORDLGLKEKGLPLALLPLRGAAESG---GAAQP-----T 472
DB AQAQAGETLDSLLEPALSFDAPLESLE--PLPALEFPDGSAREQLVLDALDPLDVA 777
QY 473 RTKAAKGV-----ELPACPCRHVDVSOAPNTGVPVQAQPAKSWDPNQLNAH 516
DB LPEAGEVSAWEGSSLEELDLSLDLPLVQLPEAEAEAPPAEALASEAPALSIAEVMA 837
QY 517 PLEPVLRGLKTRAGALRPPP 536
DB PWQPINPPAQNVPVSLPPP 857

RESULT 12

T42737
gp330 protein precursor - rat

N:Alternate names: megalin
C:Species: Rattus norvegicus (Norway rat)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C:Accession: T42737
R:Saigo, A.; Pietromonaco, S.; Loo, A.K.C.; Farquhar, M.G.
Proc. Natl. Acad. Sci. U.S.A. 91, 9725-9729, 1994
A:Title: Complete cloning and sequencing of rat gp330/megalin, a distinctive member of
A:Reference number: A58173; MUID:95024033; PMID:7937880
A:Accession: T42737
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4660 <SAI>
A:Cross-references: UNIPROT:P98158; UNIPARC:UPI000013C4B4; EMBL:L34049; NID:G561852; P
A:Experimental source: strain Sprague-Dawley; kidney
C:Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-bindi
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-4660/Product: gp330 protein #status predicted <NAT>

Query Match 4.3%; Score 124; DB 2; Length 4660;
Best Local Similarity 19.7%; Pred. No. 5.6;
Matches 118; Conservative 69; Mismatches 191; Indels 222; Gaps 31;

QY 66 YLGQLEH-EDITDSADAVEDLTEAEWEDLTQQYYSLVHGDAFISNSRNYFSQCOALLNR 123
DB YLEEEHITQIDYDWD-----PEHIGLSVYVYTVLAQSQSGFGAIKRAYIPNFESGSN- 4130
QY 124 ITSVPNQTDID-GLRNIWIIPK---AAKSRGRDIVCMDEVEIILSLAAADHPLSRDNKVV 179
DB ---NPIREVDLGLK--YLMQPDGLAVDWGHHIYWSDAKSORIEVATLD---GRYRKWL 4181
QY 180 VQKIETPLLICDTKFDIROQWFLVTDWNPILTIFWYKESYLRFSTORFSLDKLSAIIHLCN 239
DB ITTQLDQPAATA-----VNP-----KLGMEFWTQGGKPKIESA----- 4215
QY 240 NAVQKYLKNDVGRSPLLPAHNNWTSRQEYLORQGRGAVGVSIVYPSMKKAIHAHMK-- 297
DB ---WMNGEHSVLVSENGLWPNGLSIDYLDND-----RVYWSDSKEDVIEAIKYD 4262
QY 298 ---VAQDHVEP-----RKNSFELYGADFVLRDRFPWLIRI 330
DB GTDRLIINEAMKPSLDIFEDKLYVVAKEGEVWRQNKFGKENKEVL--VVPMLTVQ 4320
QY 331 N-----SSPTMHPSTPTVTAQL-----CAQVQEDTIKVAVDVRSCLDIGNFELLMRQ 374
DB RIFHLRYNQSVNSPKQVCSHLCLLRPGGYSCACPGQSDFTVGTSTVQCDAAAS-EL---- 4375
QY 375 PVVEPPP-----FSGSDLGVAGSVRRARRQVLP----- 403
DB PVTWPPPCRMHGGNCYFDENELPKCKCSGYSYG-EYCEVGLS-----RGIPPGTTMAVL 4429
QY 404 -----VCNKLKASALLDAQPLKARGPSAMPDPAQG----- 433
DB LTFVIVIVGALVLVGLFHYRKTGSLTLP-LKPLSLSLAKPSENGNGVTFRSGADVNM 4488
QY 434 -----PPSPALQRDLGLKE-----EKG-LPLALLAPLRGAESGAAQPTRTKAAGKVE 481
DB DIGVSPFPGETIIDRSMAHNEHFVMEVGKQVPIPENPMYAAKN-----TSKVA 4537
QY 482 LPACPCRHVDVSOAPNTGVPVQAQPAKSWDPNQLNAHPLFVLRLGLKTAEGALRPPPGKGS 541
DB ---QGPSTGAQVTPENVV--ENQNYGRPIDP-----SEIVPEKFPASPQA 4580

RESULT 13

A70507
probable reductase (EC 1.3.99.-) iron-sulfur protein - Mycobacterium tuberculosis (str
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C:Accession: A70507
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998

A;Authors: Sqaers, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: A70507
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-882 <COL>
A;Cross-references: UNIPROT:O33268; UNIPARC:UPI000000D4F63; GB:297991; GB:AL123456; NID:9
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0338c
C;Keywords: oxidoreductase

Query Match 4.2%; Score 120; DB 2; Length 882;
Best Local Similarity 27.2%; Pred. No. 0.98;
Matches 43; Conservative 11; Mismatches 52; Indels 52; Gaps 5;

QY 416 AQLKARGPSAMPD-----PAQGPPSPALQDGLGKEKGLPLALLAPLRGAES 465
DB 736 AKQAQERAPKAAAPVTPVPEAPAEPOAPA-----PAAPAPVKGIGMA 782
QY 466 GGAAQPTRTKAAGKVELPACPCRHVS-----QAPNTGVPVPAQPAKSW 508
DB 783 AGAKRFGCAKAAPTPAAPAPAPVKGGLGTAAGKRPAGKTPPPAPGLAEPAAPQPEA 842
QY 509 DPNOLNAHLEP-----VLRLKTAEGALRPP 535
DB 843 KPQPEAAPPKPQTGDGPAAPAPVKGGLGIARGA-RPP 879

RESULT 14
S55659
segment protein 64 - equine herpesvirus 2
C;Species: equine herpesvirus 2
C;Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 09-Jul-2004
C;Accession: S55659
R;Reiford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.
J. Mol. Biol. 249, 520-528, 1995
A;Title: The DNA sequence of equine herpesvirus 2.
A;Reference number: S55594; MUID:95302501; PMID:7783207
A;Accession: S55659
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-3436 <TEL>
A;Cross-references: UNIPROT:Q66666; UNIPARC:UPI000000F7BF4; GB:U20824; NID:96951172; PIDN:
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

Query Match 4.2%; Score 119.5; DB 2; Length 3436;
Best Local Similarity 18.6%; Pred. No. 7.8; Mismatches 261; Indels 233; Gaps 29;
Matches 134; Conservative 94;

QY 9 VVHQSCSRSSRSPRQREAGS-----SDLSSRQDAENAEAKRLGQLQGLVDIAKVC 63
DB 1919 VAATGSLPRLRSHSPAGAE DTGAALFALSEAVKLEKSGRIRWRGVEGKLSDAHSLVA 1978
QY 64 QAYLGQLEHEDI-----DTSADAVEDLTEAWEDLTQQYSLVHGDFAFTSNRNYF 114
DB 1979 RSAAKGAQKRLFALIQGLRKDYAAQREIME-DWK-----AFVT----- 2018
QY 115 SQCCALLNRITSVNPQTDIDGLRNWIIPKAASRGDIVCMRVESILELAADHPLSR 174
DB 2019 ---EAPINSMEDVND-----ILRAAPNEAEFEAKLEKEVRREAEAKLEK 2064
QY 175 DNKVVVOKYIETPLLIC-----DTKFDIRQWFLVT--DWNPLTIWFKESYLRFSSTQSFSL 228
DB 2065 EAEEMLTQAVKRGLOCGWRQIONAFDNMAFGIITGEDWAAVAAEFQREGSLTSLTPGQL 2124
QY 229 KDLDSATH-----LCNNNAVOKYLNVDVGRSP-----LLP--AH-NMWTSTRFQYLRQG 275
DB 2125 SKLTDKVEAEAEALLNKVSMPLNGPAFPKPPAFDPLTPVRAHNFYKLSFPPLKLNROA 2184
QY 276 RG-----AVWGSVIYPS-----MKKAIAHAMKV---AODHVEPRKNSFELYGADF 317

DB 2185 EAVEAKMSQIEQAIEGADVVEAVAGTPLEAPVARALLRLAARDEAAGLKQIDEGERYA 2244
QY 318 VLG-----PDRFWLIE----- 329
DB 2245 VQGVVERAGSGBPAPKPAEIPKKLLTYEQTLISLANLPEDFQKNVLQNETLMNLQUREYL 2304
QY 330 -----INSSPTWHPST--PVTAQLCAQVOEDTIKVAV-----DRSCDIGNFELLWR- 373
DB 2305 GRVTENINLSERAKTSRGEANRLAAVTEENLPQANVISISSRPLDKSDPVGFLGIVRD 2364
QY 374 QPVVEPPFSGSDLCVAGSVRRARQVLPVC--NLKASASLLDQAQPLKARGSAMPDPA 431
DB 2365 KOIVESDPYSATRESL--VMLHRTFKALLPLCPASLKRRLMELIGEBILREKG-----RG 2416
QY 432 QPPSPALQ-----RDGLKE- 447
DB 2417 RGAIPRYGVRAHETDDVGVLTAEIETLEARRVTGGKAAVEGWVKKRDAYRNWVEDLALRSE 2476
QY 448 -EKGL-PL-----ALLAPLRGAESGGAQPTRTKAAGKVELPACPCRH----- 489
DB 2477 VEKRGPLVEKSGREALDAPDLAVLQBEAAALLLEAKTGGLDK--SAPETHERVLEIQMYL 2534
QY 490 -----VDSQAPNTGVPVPAQPAKSWDPNQLNAHLEPVLRLGLKTAEGALRPPPGK 539
DB 2535 RFKLDPLKHYLDSQRPVFEAAPLSRALYWSNSQGERESRENSGEGEGVEGERRGEGGE 2594
QY 540 GS 541
DB 2595 GS 2596

RESULT 15
T41104
probable transcription factor subunit, TPR domain - fission yeast (Schizosaccharomyce
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T41104
R;Purnelle, B.; Goffeau, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21964
A;Accession: T41104
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1006 <PUR>
A;Cross-references: UNIPROT:O74458; UNIPARC:UPI000006B686; EMBL:AL031535; PIDN:CAA20753
A;Experimental source: strain 972h-; cosmid c16C4
C;Genetics:
A;Gene: SPDB:SPCC16C4.14c
A;Map position: 3

Query Match 4.0%; Score 115.5; DB 2; Length 1006;
Best Local Similarity 21.0%; Pred. No. 2.6;
Matches 92; Conservative 58; Mismatches 137; Indels 151; Gaps 24;

QY 28 EEAQSDLSRQD-----AENAEAKRLGPG--OLVDIACKVCOA---VLQLEHEDI 75
DB 3 QNGNSIVDSNMNETQNTDNDFAEQDNLNGYIIEIVDEARNVSEVDKFLG----- 55
QY 76 DTSADAVEDLTEAWEDLTQQYSLVHGDFAFTSNRNYFSQCCALLNRITSVNPQTDI-- 133
DB 56 DTSALQAEGL---WSDESDYEG-----SDDESNSFK-----TASRTEDDIAN 95
QY 134 -----DGLRNWIIPKAASRGR-----DIVCMRVEEIILELAADHPLSRDNKW- 178
DB 96 EEWEENLKA VAGFRK---VRKGHKGRVSRADMLPSVEVQOMLSL--AHNLFAQGNFD 150
QY 179 VVOKYIETPLLICDTKFDIRQWFLVTDNWPLTIW-FYKESYLRFSSTQSFSLDKL-----DS 233
DB 151 EAQKLAEEIVRI-----DNNVIAAKWMLGECHRGQRGNRVNIEKCLIAWMA 196
QY 234 AIHLCNNNAVOKYLNVDVGRSP--PAHNMW-TSTRFQYLRQGRGAVWGSVIY----- 285
DB 197 AAHL-----KP--KDHELWFTCAKLSSELE-----FWDQADYCNRAVS 233

```
QY 286 -----PSMKKATAHAKVAQDHVEPRKNSFELYGADFVLGRDRPREWLIEINSPTMHPS 339
Db :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
234 AKPPNKSLLKYYIWNRSVLNKEHGSILKAA---EGFKLL-----QS 272

QY 340 TPVTAOLCAQVOEDTIKVAVDRSCDIGNFELLWRQPVVEPPFSGSDLCVAGSVRRARR 399
Db :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
273 SPYNASILKNLAEIYIKIHAPREI-LKQFEIAWKYFYQYPAPPIGNDIF----- 320

QY 400 QVLPVCNKKASALLDAQ 417
Db :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
321 -DLPTLNLYAEULLLDHQ 337
```

Search completed: April 4, 2006, 12:51:51
Job time : 24.1115 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: April 4, 2006, 12:42:35 ; Search time 172.456 Seconds
(without alignments)
1378.343 Million cell updates/sec
Title: US-10-635-977-2
Perfect score: 2854
Sequence: 1 MASSILKVVVSHQSCRSR.....LRGLKTAEGALRPPPGKGKS 541
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2443163 seqs, 439378781 residues
Total number of hits satisfying chosen parameters: 2443163
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : A_Geneseq_21.*
- 1: Geneseqp1980s.*
 - 2: Geneseqp1990s.*
 - 3: Geneseqp2000s.*
 - 4: Geneseqp2001s.*
 - 5: Geneseqp2002s.*
 - 6: Geneseqp2003as.*
 - 7: Geneseqp2003bs.*
 - 8: Geneseqp2004s.*
 - 9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2854	100.0	541	8	ADJ93358	Adj93358 Human BGS
2	2843	99.6	541	8	ADJ93365	Adj93365 Human tub
3	1560	54.7	293	8	ADJ93366	Adj93366 Human BGS
4	1299	45.5	242	8	ADJ93360	Adj93360 Human BGS
5	1043.5	36.6	744	4	AB94796	AB94796 Human pro
6	1013	35.5	488	5	AAU74334	AAU74334 Human cyt
7	995.5	34.9	399	6	ABU11512	ABU11512 Human MDD
8	969.5	34.0	362	3	AB58909	AB58909 Breast an
9	956.5	33.5	352	3	AB43005	AB43005 Human ORF
10	956.5	33.5	352	4	AM39450	AM39450 Human pol
11	956.5	33.5	352	8	ADJ93457	Adj93457 Human ROT
12	956.5	33.5	352	8	ABM80420	ABM80420 Tumour-as
13	944.5	33.1	326	7	ADM05524	ADM05524 Human pro
14	865.5	30.3	292	8	ADJ93455	Adj93455 Human ROT
15	722.5	25.3	992	4	AB65645	AB65645 Drosophil
16	715	25.1	432	6	ABU92048	ABU92048 Human pro
17	536.5	18.8	719	4	AB65541	AB65541 Drosophil
18	524.5	18.4	281	8	ADQ66614	ADQ66614 Novel hum
19	519	18.2	362	4	AAW41236	AAW41236 Human pol
20	516	18.1	402	8	ADJ02747	Adj02747 Novel hum
21	375.5	13.2	566	4	ABG05971	ABG05971 Novel hum
22	336	11.8	160	8	ADH45424	ADH45424 Human mol
23	330	11.6	347	6	ABU00150	ABU00150 Human nov
24	328	11.5	496	4	ABB64074	ABB64074 Drosophil

25	311	10.9	673	8	ADQ66481	Adq66481 Novel hum
26	303.5	10.6	1281	4	AAW39105	AAW39105 Human pol
27	303.5	10.6	1304	4	AAW40891	AAW40891 Human pol
28	299.5	10.5	423	5	ABP43750	ABP43750 Ligase TT
29	299.5	10.5	423	8	ADJ93458	Adj93458 Human tub
30	299	10.5	393	7	ADC31411	ADC31411 Human nov
31	297.5	10.4	827	4	ABB60840	ABB60840 Drosophil
32	291	10.2	989	4	ABB60296	ABB60296 Drosophil
33	282.5	9.9	487	4	ABBS9096	ABBS9096 Drosophil
34	280.5	9.8	439	7	ADB65504	ADB65504 Human pro
35	278	9.7	321	7	ADC33066	ADC33066 Human nov
36	274	9.6	49	8	ADJ93373	Adj93373 Human BGS
37	271.5	9.5	519	4	AAW79305	AAW79305 Human pro
38	271.5	9.5	524	4	AB22915	AB22915 Protein #
39	271.5	9.5	592	7	ADB65217	ADB65217 Human pro
40	270	9.5	917	4	AB964837	AB964837 Drosophil
41	259	9.1	379	8	ADJ93456	Adj93456 Pig tubul
42	255.5	9.0	1138	8	ADJ71954	Adj71954 Human PMM
43	254.5	8.9	377	4	AAW79068	AAW79068 Human cyt
44	254.5	8.9	377	5	ABB83472	ABB83472 Human pro
45	254.5	8.9	377	7	ADB47756	ADB47756 Human NOV

ALIGNMENTS

RESULT 1
ADJ93358
ID ADJ93358 standard; protein; 541 AA.
XX AC ADJ93358;
XX DT 06-MAY-2004 (first entry)
XX DE Human BGS-42 protein sequence SeqID2.
XX KW testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytostatic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; neotropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human.
KW Homo sapiens.
XX WO2004005487-A2.
XX 15-JAN-2004.
XX 09-JUL-2003; 2003WO-US021605.
XX 09-JUL-2002; 2002US-0394725P.
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX Feder JN, Wu S, Nelson TC;
XX WPI; 2004-099381/10.
XX N-PSDB; ADJ93357.
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide, e.g.
XX useful for preventing, treating or ameliorating a medical condition, e.g.
XX aberrant cellular proliferation, reproductive disorders or testicular
XX disorders.

PS	Claim 5; SEQ ID NO 2; 343pp; English.	
XX	This invention relates to a novel testis-specific tubulin tyrosine-ligase	
CC	-like polypeptide, designated the BGS-42 polypeptide. The invention may	
CC	be useful for the development of compounds with a cytosstatic, respiratory	
CC	-Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,	
CC	antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,	
CC	antiparkinsonian, antiarthritic, antiaesthatic, anti-HIV, antibacterial,	
CC	immunosuppressive, antiseborrheic or dermatological activity acting as	
CC	tyrosine ligase modulators. In addition, the disclosed sequences may be	
CC	useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be	
CC	used for diagnosing a pathological condition or a susceptibility to a	
CC	pathological condition in a subject, and for preventing, treating or	
CC	ameliorating a medical condition, such as a disorder related to aberrant	
CC	tubulin ligase activity, a disorder related to aberrant tubulin-	
CC	carboxypeptidase activity, aberrant cellular proliferation, reproductive	
CC	disorders, testicular disorders, testicular cancer, pulmonary disorders,	
CC	lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,	
CC	neural disorders, brain cancer, liver cancer, or proliferative condition	
CC	of the testis, lung, small intestine, brain or lymph tissue. The BGS-42	
CC	polypeptide, polynucleotide, or their modulators are also useful for	
CC	treating infertility, Cushing's syndrome, emphysema, Addison's disease,	
CC	-42 polypeptide can be used as a preventive agent for immunological	
CC	diseases including arthritis, asthma, AIDS, sepsis, acne, Sjogren's	
CC	disease or scleroderma. The antibodies may be used to purify, detect and	
CC	target the BGS-42 polypeptides. The present sequence is that of the human	
CC	BGS-42 protein of the invention.	
XX	Sequence 541 AA;	
QY	Query Match	
DB	Best Local Similarity 100.0%; Score 2854; DB 8; Length 541;	
DB	Matches 541; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MASSILKVVVSHSCSSRSSKRPDQREAGSDLSRRQDAENAEAKLRLPGQLVDIAC 60	
DB	1 MASSILKVVVSHSCSSRSSKRPDQREAGSDLSRRQDAENAEAKLRLPGQLVDIAC 60	
QY	61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQQYSLVHGDAFTSNRNYFSQCOAL 120	
DB	61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQQYSLVHGDAFTSNRNYFSQCOAL 120	
QY	121 LNRITSNVPTDIDGLRNIMIIPAKSKRGDIIVCMRVREIELELAADHPLSRDNKVV 180	
DB	121 LNRITSNVPTDIDGLRNIMIIPAKSKRGDIIVCMRVREIELELAADHPLSRDNKVV 180	
QY	181 QKYIETPLLCIDTKFDIRQWFLVTDWNPFLTIFWKESYLFSTQRFSLDKLDSAIHLCNN 240	
DB	181 QKYIETPLLCIDTKFDIRQWFLVTDWNPFLTIFWKESYLFSTQRFSLDKLDSAIHLCNN 240	
QY	241 AVQKYLKNDVGRSPLPAHNWNTSTRFQYLRQGRGAVNGSVTIYPSMKKAIAHAMKVAQ 300	
DB	241 AVQKYLKNDVGRSPLPAHNWNTSTRFQYLRQGRGAVNGSVTIYPSMKKAIAHAMKVAQ 300	
QY	301 DHVEPRKNSFELYGADFVLGRDPRPWLIEINSSPTMPSTPVTQAQLCAQVOEDTIKVAVD 360	
DB	301 DHVEPRKNSFELYGADFVLGRDPRPWLIEINSSPTMPSTPVTQAQLCAQVOEDTIKVAVD 360	
QY	361 RSCDIGNFELLWRQPVVPEPPFSGSDICVAGSVRRARROVLPVCNLKASASLLDAOPLK 420	
DB	361 RSCDIGNFELLWRQPVVPEPPFSGSDICVAGSVRRARROVLPVCNLKASASLLDAOPLK 420	
QY	421 ARGPSAMPDPAQGPSPALQRDGLGKEKGLPLALLAPLRGAESGGAQAQTRTKAAGV 480	
DB	421 ARGPSAMPDPAQGPSPALQRDGLGKEKGLPLALLAPLRGAESGGAQAQTRTKAAGV 480	
QY	481 ELPACPCRHDVDSQAPNTGVPVQAQPAKSWDPNQLNAHPLPVLIRGLKTAEGALRPPPGKG 540	
DB	481 ELPACPCRHDVDSQAPNTGVPVQAQPAKSWDPNQLNAHPLPVLIRGLKTAEGALRPPPGKG 540	
QY	541 S 541	

Db	541 S 541	
RESULT 2		
ADJ93365		
ID	ADJ93365 standard; protein; 541 AA.	
XX		
AC	ADJ93365;	
XX		
DT	06-MAY-2004 (first entry)	
XX		
DE	Human tubulin tyrosine ligase protein consensus sequence SeqID13.	
XX		
KW	testis-specific tubulin tyrosine-ligase-like polypeptide;	
KW	BGS-42 polypeptide; cytosstatic; respiratory-Gen; gastrointestinal-Gen;	
KW	neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;	
KW	osteopathic; nootropic; antiparkinsonian; antiarthritic; antiaesthatic;	
KW	anti-HIV; antibacterial; immunosuppressive; antiseborrheic;	
KW	dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;	
KW	tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;	
KW	testicular disorder; testicular cancer; pulmonary disorder; lung cancer;	
KW	gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;	
KW	brain cancer; liver cancer; proliferative condition; testis; lung;	
KW	small intestine; brain; lymph tissue; infertility; Cushing's syndrome;	
KW	emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;	
KW	Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;	
KW	sepsis; acne; Sjogren's disease; scleroderma; human.	
OS	Homo sapiens.	
OS	Synthetic.	
XX		
PN	W02004005487-A2.	
XX		
PD	15-JAN-2004.	
XX		
PF	09-JUL-2003; 2003WO-US021605.	
XX		
PR	09-JUL-2002; 2002US-0394725P.	
XX		
PA	(BRIM) BRISTOL-MYERS SQUIBB CO.	
XX		
PI	Feder JN, Wu S, Nelson TC;	
XX		
DR	WPI; 2004-099381/10.	
DR	N-PSDB; ADJ93364.	
XX		
PT	New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,	
PT	useful for preventing, treating or ameliorating a medical condition, e.g.	
PT	aberrant cellular proliferation, reproductive disorders or testicular	
PT	disorders.	
XX		
PS	Example 4; SEQ ID NO 13; 343pp; English.	
XX		
CC	This invention relates to a novel testis-specific tubulin tyrosine-ligase	
CC	-like polypeptide, designated the BGS-42 polypeptide. The invention may	
CC	be useful for the development of compounds with a cytosstatic, respiratory	
CC	-Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,	
CC	antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,	
CC	antiparkinsonian, antiarthritic, antiaesthatic, anti-HIV, antibacterial,	
CC	immunosuppressive, antiseborrheic or dermatological activity acting as	
CC	tyrosine ligase modulators. In addition, the disclosed sequences may be	
CC	useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be	
CC	used for diagnosing a pathological condition or a susceptibility to a	
CC	pathological condition in a subject, and for preventing, treating or	
CC	ameliorating a medical condition, such as a disorder related to aberrant	
CC	tubulin ligase activity, a disorder related to aberrant tubulin-	
CC	carboxypeptidase activity, aberrant cellular proliferation, reproductive	
CC	disorders, testicular disorders, testicular cancer, pulmonary disorders,	
CC	lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,	
CC	neural disorders, brain cancer, liver cancer, or proliferative condition	
CC	of the testis, lung, small intestine, brain or lymph tissue. The BGS-42	
CC	polypeptide, polynucleotide, or their modulators are also useful for	
CC	treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's	

Db 121 TKFDIRQWFLVTDNPLTIWFKYKESYLRFSTQFSLDKLDSAHLHCNNAVQKYLKNDVGR 180
Qy 253 SPLLPAHNNMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
Db 181 SPLLPAHNNMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 240
Qy 313 YGADFVLGRDPRFLWLEINSSPTMHPSTPVTALCAQVQEDTIKVAVDKSCDI 365
Db 241 YGADFVLGRDPRFLWLEINSSPTMHPSTPVTALCAQVQEDTIKVAVDKSCDI 293

RESULT 4

ADJ93360

ID ADJ93360 standard; protein; 242 AA.

XX AC ADJ93360;

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of the human
CC BGS-42 protein (partial sequence) of the invention.
XX

Qy Sequence 242 AA;

Query Match 45.5%; Score 1299; DB 8; Length 242;

Best Local Similarity 100.0%; Pred. No. 1.7e-115;

Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 133 IDGLRNIIWKPAKSRGRDVCMDRVEILEAAADHPLSRDNKVVQKIETPLLCD 192

Db 1 IDGLRNIIWKPAKSRGRDVCMDRVEILEAAADHPLSRDNKVVQKIETPLLCD 60

Qy 193 TKFDIRQWFLVTDNPLTIWFKYKESYLRFSTQFSLDKLDSAHLHCNNAVQKYLKNDVGR 252

Db 61 TKFDIRQWFLVTDNPLTIWFKYKESYLRFSTQFSLDKLDSAHLHCNNAVQKYLKNDVGR 120

Qy 253 SPLLPAHNNMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312

Db 121 SPLLPAHNNMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180

Qy 313 YGADFVLGRDPRFLWLEINSSPTMHPSTPVTALCAQVQEDTIKVAVDKSCDI 372

Db 181 YGADFVLGRDPRFLWLEINSSPTMHPSTPVTALCAQVQEDTIKVAVDKSCDI 240

Qy 373 RQ 374

Db 241 RQ 242

RESULT 5

AAB94796

ID AAB94796 standard; protein; 744 AA.

XX AC AAB94796;

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

Db 121 TKFDIRQWFLVTDNPLTIWFKYKESYLRFSTQFSLDKLDSAHLHCNNAVQKYLKNDVGR 180
Qy 253 SPLLPAHNNMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
Db 181 SPLLPAHNNMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 240
Qy 313 YGADFVLGRDPRFLWLEINSSPTMHPSTPVTALCAQVQEDTIKVAVDKSCDI 365
Db 241 YGADFVLGRDPRFLWLEINSSPTMHPSTPVTALCAQVQEDTIKVAVDKSCDI 293

RESULT 4

ADJ93360

ID ADJ93360 standard; protein; 242 AA.

XX AC ADJ93360;

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

XX DT 06-MAY-2004 (first entry)

CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of the human
CC BGS-42 protein (partial sequence) of the invention.
XX

Qy Sequence 242 AA;

Query Match 45.5%; Score 1299; DB 8; Length 242;

Best Local Similarity 100.0%; Pred. No. 1.7e-115;

Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 133 IDGLRNIIWKPAKSRGRDVCMDRVEILEAAADHPLSRDNKVVQKIETPLLCD 192

Db 1 IDGLRNIIWKPAKSRGRDVCMDRVEILEAAADHPLSRDNKVVQKIETPLLCD 60

Qy 193 TKFDIRQWFLVTDNPLTIWFKYKESYLRFSTQFSLDKLDSAHLHCNNAVQKYLKNDVGR 252

Db 61 TKFDIRQWFLVTDNPLTIWFKYKESYLRFSTQFSLDKLDSAHLHCNNAVQKYLKNDVGR 120

Qy 253 SPLLPAHNNMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312

Db 121 SPLLPAHNNMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180

Qy 313 YGADFVLGRDPRFLWLEINSSPTMHPSTPVTALCAQVQEDTIKVAVDKSCDI 372

Db 181 YGADFVLGRDPRFLWLEINSSPTMHPSTPVTALCAQVQEDTIKVAVDKSCDI 240

Qy 373 RQ 374

Db 241 RQ 242

RESULT 5

AAB94796

ID AAB94796 standard; protein; 744 AA.

XX AC AAB94796;

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

XX DT 26-JUN-2001 (first entry)

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX Claim 8; SEQ ID NO 15921; 2537bp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 744 AA;

Query Match 36.6%; Score 1043.5; DB 4; Length 744;
Best Local Similarity 40.5%; Pred. No. 3.2e-90;
Matches 234; Conservative 85; Mismatches 192; Indels 67; Gaps 12;

QY 2 ASSILKVVVSHOSCRSRKPRDQREAGSDLSRQDAENAEAKLGLPQLVDIACK 61
DB 151 ARNLVLKVVKSE-----WKSYPIQAVEEASGDKQPKQKNPVL-----VSPFEFDEALC 201
QY 62 VQOAVLGOLEHEDIDTSADAVEDLTAEWEDLTQYVSLVGDFAFISNRYFSOCQALL 121
DB 202 ACEEVLNLAHWDIDKLEAPLYLTPEGWSLFLQYVQVHGBAELRLDVTQVQCEDIL 261
QY 122 NRITSVNPQTDIDGLNIIWIKPAKSRGRDVCMDRVEEILELAADHPLSRDNKVVQ 181
DB 262 QQLQAVVQIDMEGRNIWIKPGAKSRGRGIMCWDHLEMLKLVNGPNVVMKDGKVVQ 321
QY 182 KYIETPLICDTKFDIRQWFLVTDNPLTIFPKESYLRFTQRFSLDKLDSAIHLNNA 241
DB 322 KYIERPLIFGTGKFLQWFLVTDNPLTIFVYRDSYIRFTQPFSLNLDNSVHLNNS 381
QY 242 VQYLKNDVGRSPILPAHNMWITSTFOEVLQQRGAVMGSVIYPSMKKATAHAMKVAQD 301
DB 382 IQKLENSCHRRHPLDPNNMSSQRFQHLQDMGAPNAWSTIIIVPGMDAVIHALQTSQD 441
QY 302 HVEPRKNSFELYGDFVLGRDPRPWLIIINSPTMHPSTPTVAQLCAQVQEDTIKVAV-- 359
DB 442 TVQCRKASFELYGADFVFGDFQFWLIEINASPTWAPSTAVTARLCAGVQADTLRVVIDR 501
QY 360 ---DRSCDIGNFELLWRQVPPPPPSGSDLCVAGSVVRRARRQVLPCVNLKASALLDAQ 417
DB 502 MLDNRCDTGAFELIYKQPAVEVPQVVGIRLIVEGTIKK-----PMAMCHRRMGRVDAV 555
QY 418 PLKARGPS-----AMPDPAQGPSPALORDLGL-----KEEGK 450
DB 556 PLLTQRGSGEARHPPSLHTKQAL-LPSPHVLRHQGVLRROHSLKLVGFKALSTTGKALRT 614
QY 451 LPLA----LLAPLRGAESGGAQPTTKA-----AGKVELPACPCRHVDSQ--AP-NTG 498
DB 615 LPTAKVFLISLPNLDKFAVPSILKPKAPALLCLRGFPQLEVPCCCLPKSEQFLAPVGRS 674
QY 499 VFVAQPAKSWDPNQLNAPHLPEFVLRLGLKTAEGALRPPP 536

DB 675 RPKANRRPDCDKPRAEACPMKRL-----SPLKPLP 704

RESULT 6
AAU74334

ID AAU74334 standard; protein; 488 AA.

XX AAU74334;

DT 12-MAR-2002 (first entry)

DE Human cytoskeleton-associated protein (CYSKP) #5.

XX Human; cytoskeleton-associated protein; CYSKP; autoimmune disorder;
KW cell proliferative disorder; inflammatory disorder; prion disease;
KW vesicle trafficking disorder; gastrointestinal disorder; muscle disorder;
KW neurological disorder; cell motility disorder; reproductive disorder;
KW spinal cord disease; central nervous system disorder; mental disorder;
KW gene therapy; cancer.

XX Homo sapiens.

XX WO200185942-A2.

XX 15-NOV-2001.

XX 03-MAY-2001; 2001WO-US014355.

XX 05-MAY-2000; 2000US-0201960P.

XX 08-MAY-2000; 2000US-0202729P.

XX 05-JUN-2000; 2000US-0209705P.

XX 07-JUN-2000; 2000US-0210149P.

XX 21-JUN-2000; 2000US-0213215P.

XX (INCY-) INCYTE GENOMICS INC.

XX Yue H, Tang YT, Au-Young J, Lu DAM, Baughn MR, Hillman JL;
PI Azimzai Y, Lal P, Yao MG, Bandman O, Burford N, Battra S, Kearney L;
PI Policky JL;

XX WPI; 2002-062248/08.
DR N-PSDB; AAS99894.

XX New cytoskeleton-associated proteins and polynucleotides, useful for
PT diagnosing, preventing and treating cell proliferative, autoimmune,
PT inflammatory, neurological, cell motility, reproductive and muscle
PT disorders.

XX Claim 1; Page 130-131; 194pp; English.

XX The invention relates to human cytoskeleton-associated polypeptides
CC (CYSKP) and their associated polynucleotide sequences. The sequences are
CC useful in the treatment of disorders associated with overexpression or
CC underexpression of CYSKP in a patient. The disorders include cell
CC proliferative disorders (such as cancer, actinic keratosis,
CC arteriosclerosis, cirrhosis, hepatitis and psoriasis),
CC autoimmune/inflammatory disorders (such as, asthma, atherosclerosis,
CC osteoporosis, Crohn's disease, rheumatoid arthritis, diabetes mellitus
CC and anaemia), vesicle trafficking disorders (such as
CC hypercholesterolaemia, diabetes insipidus, Grave's disease and goitre),
CC gastrointestinal disorders, prion diseases, neurological disorders (such
CC as epilepsy, stroke, cerebral neoplasms, Alzheimer's disease,
CC Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis
CC and other motor neuron disorders), cell motility disorders, reproductive
CC disorders (such as endometriosis and polycystic ovary syndrome), muscle
CC disorders (such as myocarditis, migraine, hypertension, hypoglycaemia,
CC myocardial infarction, epilepsy and muscular dystrophy), spinal cord
CC diseases, central nervous system disorders (such as Down syndrome and
CC cerebral palsy) and mental disorders (such as anxiety and schizophrenia).

XX Sequences AAU74330-AAU74363 represent human CYSKP of the invention
XX Sequence 488 AA;

PR	20-JUN-2001; 2001US-0300001P.
XX	(INCY-) INCYTE GENOMICS INC.
XX	Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI	Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshay SR;
PI	Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
PI	Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI	Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX	WPI: 2003-058431/05.
DR	N-PSDB; ABX34502.
XX	New purified disease detection and treatment molecule proteins and polynucleotides, useful for diagnosing, treating or preventing cancers (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis or hepatitis.
XX	Claim 27; SEQ ID NO 459; 339pp + Sequence Listing; English.
XX	This invention describes a novel disease detection and treatment molecule polypeptide (MDTT) which has anti-inflammatory, immunosuppressive, osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic, antianaemic, antipsoriatic and hepatotropic activity. The polynucleotides and the polypeptides of the invention can be used for gene therapy, protein replacement therapy and are useful for treating a variety of diseases or conditions. These polypeptides or polynucleotides are particularly useful for diagnosing, treating or preventing cell proliferative disorders (e.g. cancers including adenocarcinoma, leukaemia, lymphoma, melanoma or sarcoma), anaemia, Crohn's disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's syndromes, inflammation, osteoporosis, thrombocytopaenia, psoriasis or hepatitis. ABU11450-ABU11845 represent the MDTT polynucleotides encoded by ABU11450-ABU11845, described in the disclosure of the invention. NOTE: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	Sequence 399 AA;
SQ	
Query Match	34.9%; Score 995.5; DB 6; Length 399;
Best Local Similarity	53.4%; Pred. No. 4.9e-86;
Matches	187; Conservative 57; Mismatches 101; Indels 5; Gaps 2;
QY	54 QLVDIACKVQAVLGQLEHEDITSADAVEDLTEAEWEDLTQOYSLVHGDAFISNRNY 113
DB	29 EFVDEALCACEYLSNLAHMDIDKLEAPLYLTPEGWSLFQRYQVYVHEGAEIRHLDQ 88
QY	114 FSCQALLNRITSVNPQTDIDGLNRNIWIIPKAASRGRDITVCMRVEEILELAADHPLS 173
DB	89 VQCEDILOQLQAVVQIDMEGRNIWIVKPGAKSRGRGIMCMCHLEMLKLVNGPNVVM 148
QY	174 RDNKVVQKYIETPLICTDKFDIRQWFLVTDWNPLTITWYKESYIRFSTQRFSLDKLDS 233
DB	149 KDGVVQKYIETPLIFGTKFDLRQWFLVTDWNPLTITWYKESYIRFSTQRFSLDKLDS 208
QY	234 AIHLCHNNAVQYKLVNLDVGRSPVLLPAHNMWTSFQOYVLRQGRGAVGWSVIYPSMKKAIA 293
DB	209 SVHLCHNNSIQKHLNCHRHPLLPDNNMWSQRFQAHLOEMGAPNAWSTIIVPGMKDAVI 268
QY	294 HAMKVAQDHVEPRKNSFELYGADFVLGRDPRPWLIEINSSPTMHPSTVPTAQLCAOVQED 353
DB	269 HALQTSQDVTQCRKASFELYGADFVGEDFQPWLIENASFTWAPSTAVTARLCAGVQAD 328
QY	354 TIKVAV----DRSCDIGNFELLRQPVVEPPFSGSDLCVAGV-SVRRAR 398
DB	329 TLRWIDRLDRNCDTGAPELIVKQPVTTSTPASTPRSPCLLPMTYSDTRAR 378
RESULT 8	
AAB58909	
ID	AAB58909 standard; protein; 362 AA.
XX	

CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antithematic; antithyroid; and antianaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ Sequence 352 AA;

Query Match 33.5%; Score 956.5; DB 3; Length 352;
Best Local Similarity 54.2%; Pred. No. 2.2e-82;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;

QY 74 DIDTSADAVEDLTEAEWEDLTQYYSLVHGDAFISNRNYSFQOALLNRITSVNPQTDI 133
Db 2 DIDKLEAPLYLTPEGWSLFLQRYQVWHEGAEHLDTQVQRCEDILQQLQAVVPQIDM 61

QY 134 DGLRNWIIPKAASGRDIVCMRDVEEILLEAAADHPLSRDNKNVQVQKIETPLLICDT 193
Db 62 EGRNWIWPKGAKSRGIMCMDHLEMLKLVNGPNVVMKDGKVVQVQKIETPLIFGT 121

QY 194 KFDIROWFLVDNPLTIWFKESYLRFSTQRFSLDKLSAIAHLNNVAKVQKLVNDVGRS 253
Db 122 KFDLRQWFLVDNPLTIWFRDYSIRFSTQPSLKNLDSNVHLNNSIQHLENSCHRH 181

QY 254 PLLPAHNMWTSRFOEYLQQRGAVGWSVIYPSMKKAIHAAMKVAQDVHPRKNSFELY 313
Db 182 PLLPPDNMWSQRFOAHLQEMGAPNAWSTIIVPGMKDAVIHALQTSQDTVQCRKASFELY 241

QY 314 GADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKAVV----DRSCDIGNFE 369
Db 242 GADVFGEDEFPWLIEINASPTMAPSTAVTARLCAGVQADTLRVWIDRLDRNCDTGAPE 301

QY 370 LLWRQPVVEPPFSGDLGVAGV-SVRRAR 398
Db 302 LIYQPVTTSPASTPRPSCLLPMSYDTRAR 331

RESULT 10
AAM39450
ID AAM39450 standard; protein; 352 AA.
XX
AC AAM39450;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2595.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-brager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.

PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00582317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI58606.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
XX as central nervous system injuries.
XX
PS Example 4; SEQ ID NO 2595; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-brager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 352 AA;

Query Match 33.5%; Score 956.5; DB 4; Length 352;
Best Local Similarity 54.2%; Pred. No. 2.2e-82;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;

QY 74 DIDTSADAVEDLTEAEWEDLTQYYSLVHGDAFISNRNYSFQOALLNRITSVNPQTDI 133
Db 2 DIDKLEAPLYLTPEGWSLFLQRYQVWHEGAEHLDTQVQRCEDILQQLQAVVPQIDM 61

QY 134 DGLRNWIIPKAASGRDIVCMRDVEEILLEAAADHPLSRDNKNVQVQKIETPLLICDT 193
Db 62 EGRNWIWPKGAKSRGIMCMDHLEMLKLVNGPNVVMKDGKVVQVQKIETPLIFGT 121

QY 194 KFDIROWFLVDNPLTIWFKESYLRFSTQRFSLDKLSAIAHLNNVAKVQKLVNDVGRS 253
Db 122 KFDLRQWFLVDNPLTIWFRDYSIRFSTQPSLKNLDSNVHLNNSIQHLENSCHRH 181

QY 254 PLLPAHNMWTSRFOEYLQQRGAVGWSVIYPSMKKAIHAAMKVAQDVHPRKNSFELY 313
Db 182 PLLPPDNMWSQRFOAHLQEMGAPNAWSTIIVPGMKDAVIHALQTSQDTVQCRKASFELY 241

QY 314 GADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKAVV----DRSCDIGNFE 369
Db 242 GADVFGEDEFPWLIEINASPTMAPSTAVTARLCAGVQADTLRVWIDRLDRNCDTGAPE 301

QY 370 LLWRQPVVEPPFSGDLGVAGV-SVRRAR 398
Db 302 LIYQPVTTSPASTPRPSCLLPMSYDTRAR 331

RESULT 11
ADJ93457
ID ADJ93457 standard; protein; 352 AA.

XX AC ADJ93457;
 XX DT 06-MAY-2004 (first entry)
 XX DE Human HOTT13 protein sequence SeqID2.
 XX KW testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
 KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
 KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human; HOTT13.
 XX OS Homo sapiens.
 XX PN WO2004005487-A2.
 XX XX 15-JAN-2004.
 XX PF 09-JUL-2003; 2003WO-US021605.
 XX PR 09-JUL-2002; 2002US-0394725P.
 XX PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX PI Feder JN, Wu S, Nelson TC;
 XX DR WPI; 2004-099381/10.
 XX PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
 PT useful for preventing, treating or ameliorating a medical condition, e.g.
 PT aberrant cellular proliferation, reproductive disorders or testicular
 PT disorders.
 XX PS Disclosure; SEQ ID NO 7; 343pp; English.
 XX CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
 CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytosolic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, anti-HIV, antibacterial,
 CC antiparkinsonian, antiarthritic, antiasthmatic, antiseborrheic or dermatological activity acting as
 CC immunosuppressive, antiseborrheic or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of the human
 CC HOTT13 protein which is related to the invention. Note: The present
 CC sequence does not appear in the specification but was obtained from
 CC Genbank.

XX SQ Sequence 352 AA;
 XX Query Match 33.5%; Score 956.5; DB 8; Length 352;
 XX Best Local Similarity 54.2%; Pred No. 2.2e-82;
 XX Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;
 QY 74 DIDTSADAVDELTEAWEDLTOQYYSVLVHGDAFISNSRNYFSQOALLNRITSVNPQTDI 133
 DB 2 DIDKLEAPLYLTPEGSLFLQRYQYVVHGAELRHLDTQVQRCEDILQOLQAVVPQIDM 61
 QY 134 DGLRNWIKPAKSRGRDIIWMDRVEILELAADHPLSRNKKVWQYIETPLLICDT 193
 DB 62 EGDRIWTVKPGKSRGRGIMCMHLEMLKLVNCGNVVMKDGKWKVQYIERPLLIFFT 121
 QY 194 KFDIRQWFLVTDWNPPLTWFKESYLRSTORFSLDKLDSATHLCNNAVOKYLKNDVGRS 253
 DB 122 KFDLRQWFLVTDWNPPLTWFKESYLRSTORFSLDKLDSATHLCNNAVOKYLKNDVGRS 181
 QY 254 PLLPAHNNWTSRFOEYLQROGRGAVGWSVIYPSMKKATAHAMKVAQDHVEPRKNSFELY 313
 DB 182 PLLPPDNWSSQRFQHLQEMGAPNAWSTIIIVPGMKDAVIHALQTSQDTVQCRKASFELY 241
 QY 314 GADFVLGRDPRFWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAV----DRSCDIGNFE 369
 DB 242 GADFVFGEDFQFWLIEINASSPTMAPSTAVTARLCAGVQADTLRVVVIDRMLDRNCDTGAFE 301
 QY 370 LLWRQPVVEPPFPFGSDLCVAGV-SVRRAR 398
 DB 302 LIYQPVTTSPASTPRPSCLLPMSYDTRAR 331
 RESULT 12
 ID AEM80420
 XX AC AEM80420 standard; protein; 352 AA.
 XX DT 18-NOV-2004 (first entry)
 XX DE Tumour-associated antigenic target (TAT) polypeptide PRO80956, SEQ:1058.
 XX KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
 KW tumour; diagnosis; cell proliferative disorder; breast cancer;
 KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
 KW central nervous system cancer; bladder cancer; pancreatic cancer;
 KW cervical cancer; melanoma; leukaemia; hybridisation probe;
 KW chromosome identification; chromosome mapping; gene mapping;
 KW gene therapy; cytostatic.
 XX OS Homo sapiens.
 XX PN WO2004030615-A2.
 XX XX 15-APR-2004.
 XX PF 29-SEP-2003; 2003WO-US028547.
 XX PR 02-OCT-2002; 2002US-0414971P.
 XX PA (GETH) GENENTECH INC.
 XX PI Wu TD, Zhang Z, Zhou Y;
 XX DR WPI; 2004-347921/32.
 XX DR N-PSDB; ACN37881.
 XX PT New tumor-associated antigenic target polypeptides and nucleic acids,
 PT useful in preparing a medicament for treating or detecting a
 PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
 PT prostate cancer or tumor.
 XX PS Claim 12; SEQ ID NO 1058; 7273pp; English.

XX The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT polypeptide of the invention
XX
SQ Sequence 352 AA;

Query Match 33.5%; Score 956.5; DB 8; Length 352;
Best Local Similarity 54.2%; Pred. No. 2.2e-82;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;

QY 74 DITSDADVEDLTEAEWEDLTQOYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTDI 133
DB 2 DIDKDLEAPLYLTPEGWSLFQRYQVHGAELRLDTQVQCEDILQQLQAVVPQIDM 61
QY 134 DGLRNIIWKPAKSRGRDIVCMRDVEEILELAAADHPLSRDNKVVQKYIETPLLICDT 193
DB 62 EGDNRNIWIKPAKSRGRGIMCMCHLEMLKLVNGNPVVMKDGKVVQKYIERPLLIFFT 121
QY 194 KFDIROWFLVDNPLTIWYKESYLRFSTQFSLKNDLSAHLNNAVOKYLNKDVGRS 253
DB 122 KFDLROWFLVDNPLTIWYKESYLRFSTQFSLKNDLSAHLNNAVOKYLNKDVGRS 181
QY 254 PLLPAHNMWTSRFOEYLQGRGAVGWSVIYPSMKKAIHAAMKVAQDHVEPRKNSFELY 313
DB 182 PLLPPDNMWSQRFQAHQEMGAPNAWSTIIVFGMKDAVIHALQTSQDTVQCRKASFELY 241
QY 314 GADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKAVV----DRSCDIGNFE 369
DB 242 GADVFEGEDFQPLWIEINASPTMAPSTAVTARLCAGVQADTLRVVDRMLDRNCDTGAFE 301
QY 370 LLWRQPVVEPPFSGSLCVAGV-SVRRAR 398
DB 302 LIYKQPVTTSFASPTPRPSCLLPMYSDTRAR 331

RESULT 13
ADM05524
ID ADM05524 standard; protein; 326 AA.

XX ADM05524;
AC
XX 20-MAY-2004 (first entry)
DT
XX Human protein of the invention SEQ ID NO:4209.
DE
XX human; gene therapy; diagnostic marker; pharmaceutical.
KW
XX Homo sapiens.OS
XX EP1347046-A1.
PN
XX 24-SEP-2003.
PD
XX 12-APR-2002; 2002EP-00008400.
PP
XX

PR 22-MAR-2002; 2002JP-00137785.
XX
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
PA
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-723558/69.
DR N-PSDB; ADM03081.
XX
XX New polynucleotides and polypeptides are useful in gene therapy, for
PT developing a diagnostic marker or medicines for regulating their
PT expression and activity, or as a target of gene therapy.
XX
XX Claim 1; SEQ ID NO 4209; 305pp; English.
XX
XX The invention relates to a novel human polynucleotide and the encoded
CC polypeptide. A polynucleotide of the invention may have a use in gene
CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
CC as a primer for synthesizing the polynucleotide or as a probe for
CC detecting the polynucleotide. The polynucleotides ADM0316-ADM03758 are
CC useful in gene therapy, for developing a diagnostic marker or medicines
CC for regulating their expression and activity, or as a target of gene
CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
CC are useful as pharmaceutical agents. The present sequence represents a
CC protein sequence of the invention.
XX
SQ Sequence 326 AA;

Query Match 33.1%; Score 944.5; DB 7; Length 326;
Best Local Similarity 54.3%; Pred. No. 2.8e-81;
Matches 176; Conservative 55; Mismatches 86; Indels 7; Gaps 2;

QY 74 DITSDADVEDLTEAEWEDLTQOYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTDI 133
DB 2 DIDKDLEAPLYLTPEGWSLFQRYQVHGAELRLDTQVQCEDILQQLQAVVPQIDM 61
QY 134 DGLRNIIWKPAKSRGRDIVCMRDVEEILELAAADHPLSRDNKVVQKYIETPLLICDT 193
DB 62 EGDNRNIWIKPAKSRGRGIMCMCHLEMLKLVNGNPVVMKDGKVVQKYIERPLLIFFT 121
QY 194 KFDIROWFLVDNPLTIWYKESYLRFSTQFSLKNDLSAHLNNAVOKYLNKDVGRS 253
DB 122 KFDLROWFLVDNPLTIWYKESYLRFSTQFSLKNDLSAHLNNAVOKYLNKDVGRS 181
QY 254 PLLPAHNMWTSRFOEYLQGRGAVGWSVIYPSMKKAIHAAMKVAQDHVEPRKNSFELY 313
DB 182 PLLPPDNMWSQRFQAHQEMGAPNAWSTIIVFGMKDAVIHALQTSQDTVQCRKASFELY 241
QY 314 GADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKAVV----DRSCDIGNFE 369
DB 242 GADVFEGEDFQPLWIEINASPTMAPSTAVTARLCAGVQADTLRVVDRMLDRNCDTGAFE 301
QY 370 LLWRQ---PVVEPPFSGSLCVA 390
DB 302 LIYKQPAPNMQVSPERNAPLCPA 325

RESULT 14
ADJ93455
ID ADJ93455 standard; protein; 292 AA.

XX ADJ93455;
AC
XX 06-MAY-2004 (first entry)
DT
XX Human HOTTTL protein sequence SeqID2.
DE

KW testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytostatic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2006, 09:58:51 ; Search time 4791 Seconds

(without alignments)
450.890 Million cell updates/sec

Title: US-10-635-977-2

Perfect score: 2854

Sequence: 1 MASSILKVVVSHQSCRSR.....LRGLTAEGALRPPPGKGKS 541

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5	
Fgapop 6.0	Fgapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 9263891 seqs, 1996499642 residues

Total number of hits satisfying chosen parameters: 18527782

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/abs/ABSSWEB/spool/US10635977/runat_04042006_103546_10350/app_query.fasta.1
-DB=Published Applications NA New -QFMT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNIT8=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -HOST=abs805p
-USER=US10635977 @CGN 1.1.986 @runat_04042006_103546_10350 -NCPU=6 -ICPU=3
-NO_WMAP -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRBADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA New:

1:	/SID55/ptodata/2/pubpna/US08_NEW_PUB.seq:
2:	/SID55/ptodata/2/pubpna/US06_NEW_PUB.seq:
3:	/SID55/ptodata/2/pubpna/US07_NEW_PUB.seq:
4:	/SID55/ptodata/2/pubpna/PCT_NEW_PUB.seq:
5:	/SID55/ptodata/2/pubpna/US09_NEW_PUB.seq:
6:	/SID55/ptodata/2/pubpna/US09_NEW_PUB.seq:
7:	/SID55/ptodata/2/pubpna/US10_NEW_PUB.seq:
8:	/SID55/ptodata/2/pubpna/US10_NEW_PUB.seq:
9:	/SID55/ptodata/2/pubpna/US10_NEW_PUB.seq:
10:	/SID55/ptodata/2/pubpna/US10_NEW_PUB.seq:
11:	/SID55/ptodata/2/pubpna/US11_NEW_PUB.seq:
12:	/SID55/ptodata/2/pubpna/US11_NEW_PUB.seq:
13:	/SID55/ptodata/2/pubpna/US11_NEW_PUB.seq:
14:	/SID55/ptodata/2/pubpna/US11_NEW_PUB.seq:
15:	/SID55/ptodata/2/pubpna/US60_NEW_PUB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	922	32.3	1684	8	US-10-955-054A-43 Sequence 43, Appli
2	357	12.5	666	6	US-09-925-065A-953009 Sequence 953009,

	3	302	10.6	2624	11	US-11-072-512-1688	Sequence 1688, Ap
	4	271.5	9.5	2074	11	US-11-072-512-1401	Sequence 1401, Ap
	5	210	7.4	2111	11	US-11-072-512-1651	Sequence 1651, Ap
	6	144.5	5.1	2250	9	US-10-932-182A-167	Sequence 167, App
	7	144.5	5.1	2250	9	US-10-932-182A-167	Sequence 167, App
	8	141	4.9	1319	8	US-10-750-185-41745	Sequence 41745, A
	9	141	4.9	1319	8	US-10-750-623-41745	Sequence 41745, A
	10	136.5	4.8	47460	14	US-11-124-368A-2877	Sequence 2877, Ap
	11	132.5	4.6	10373	8	US-10-821-234-64	Sequence 64, Appli
	12	127	4.4	2534	14	US-11-219-995-1	Sequence 1, Appli
	13	126.5	4.4	2754	9	US-10-330-773-662	Sequence 662, App
	14	126	4.4	5354	14	US-11-136-527-1889	Sequence 1889, Ap
	15	126	4.4	14248	8	US-10-995-561-13381	Sequence 13381, A
	16	126	4.4	96128	8	US-10-995-561-13197	Sequence 13197, A
	17	123.5	4.3	3482	14	US-11-064-246-9	Sequence 9, Appli
	18	123.5	4.3	3482	14	US-11-064-246-11	Sequence 11, Appli
	19	122.5	4.3	997	10	US-10-301-480-609833	Sequence 609833, A
	20	122.5	4.3	997	10	US-10-301-480-1223242	Sequence 1223242, A
	21	122.5	4.3	5382	9	US-10-501-035-116	Sequence 116, App
	22	121	4.2	6409	9	US-10-501-035-158	Sequence 158, App
	23	121	4.2	15457	14	US-11-136-527-2809	Sequence 2809, Ap
	24	120	4.2	551	6	US-09-925-065A-360042	Sequence 360042, A
	25	120	4.2	551	6	US-09-925-065A-360043	Sequence 360043, A
	26	119	4.2	1724	6	US-09-925-065A-552157	Sequence 552157, A
	27	119	4.2	1724	6	US-09-925-065A-552158	Sequence 552158, A
	28	119	4.2	1724	10	US-10-301-480-530107	Sequence 530107, A
	29	119	4.2	1724	10	US-10-301-480-530108	Sequence 530108, A
	30	119	4.2	1724	10	US-10-301-480-1143516	Sequence 1143516, A
	31	119	4.2	1724	10	US-10-301-480-1143517	Sequence 1143517, A
	32	118	4.1	1724	6	US-09-925-065A-552159	Sequence 552159, A
	33	118	4.1	1724	10	US-10-301-480-530109	Sequence 530109, A
	34	118	4.1	1724	10	US-10-301-480-1143518	Sequence 1143518, A
	35	117.5	4.1	2980	11	US-11-072-512-353	Sequence 353, App
	36	117.5	4.1	3261	8	US-10-624-932-7	Sequence 7, Appli
	37	117.5	4.1	3465	8	US-10-624-932-5	Sequence 5, Appli
	38	117.5	4.1	3734	8	US-10-131-826A-147	Sequence 147, App
	39	117.5	4.1	3734	9	US-10-973-115B-147	Sequence 147, App
	40	117	4.1	1263	9	US-10-016-686-14	Sequence 14, Appli
	41	117	4.1	1862	8	US-10-750-185-58437	Sequence 58437, A
	42	117	4.1	1862	8	US-10-750-623-58437	Sequence 58437, A
	43	117	4.1	2430	9	US-10-643-457-3	Sequence 3, Appli
	44	116.5	4.1	1488	14	US-11-219-995-3	Sequence 3, Appli
	45	116.5	4.1	2149	11	US-11-096-568A-14615	Sequence 14615, A

ALIGNMENTS

RESULT 1
US-10-955-054A-43
; Sequence 43, Application US/10955054A
; Publication No. US20050266420A1
; GENERAL INFORMATION:
; APPLICANT: PUSZTAI, LAJOS
; APPLICANT: SYMMANS, W. FRASER
; APPLICANT: HESS, KENNETH R.
; APPLICANT: AYERS, MARK
; APPLICANT: STEC, JAMES
; TITLE OF INVENTION: MULTIGENE PREDICTORS OF RESPONSE TO CHEMOTHERAPY
; FILE REFERENCE: UTXC:880US
; CURRENT APPLICATION NUMBER: US/10/955.054A
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 195
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
; LENGTH: 1684
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-955-054A-43

Alignment Scores: 2.37e-60 Length: 1684
Pred. No.: 922.00 Matches: 209
Score: 55.5% Conservative: 68
Percent Similarity:

[illegible]

```

Db 666 CCTCTTTCCAGGACAAAGTGGTGGTCCAGAGTACATCAGACGCGCTGCTCATC 607
Qy 191 CysAspThrLysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThr 210
Db 606 TGTGACCAAGTTCGACATCAGACAGTGGTCTCGTCACGACTGGACCCCTGACC 547
Qy 211 IleTrpPheTyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLys 230
Db 546 ATCTGGTCTTACAAGGAGAGTTACTTGGCGTTCTCAACTCAGCGCTTCTCCCTGACAA 487
Qy 231 LeuAspSerAlaIleHis-----Leu 237
Db 486 CTGGACAGTCAAGTGGTGGCGGCGACCTGGACACCTCGGCGCAGGAAATGGCTCTGCC 427
Qy 238 CysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuPro 257
Db 426 TGCATGCTGGAGCAGCCGCA---GTGAGAACGACGACGGGCGAGGCCCTCATCTGCCA 370
Qy 258 AlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGly 277
Db 369 CCT-----GCCACACGAGAGGCACAGAT-----GGGAAACT 337
Qy 278 AlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLys 297
Db 336 GAGGCTTGGGAGGCTGATGATCTCCCGGAGGACAG-----301
Qy 298 ValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPhe 317
Db 300 ---GCCACGAGCCAC-----289
Qy 318 ValLeuGlyArgAspPheArgProTrpLeuIleAsnSerSerProThrMethis 337
Db 288 -----CGGCTT-----TCAAAGCCAGCG-----271
Qy 338 ProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysVal 357
Db 270 CCTACACCCCGCTCTGTGCATCATCTCGAAATCTTAAAGGCGACCATTT---TTG 214
Qy 358 AlaValAspArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValVal 377
Db 213 CCTGTGCCAGGGG-----ATTGGA-----ATGGTGTGG-----GCA 181
Qy 378 GluProProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAla 397
Db 180 GCACCTGGGCGCTGG-----ATGGA-----ATGGTGTGG-----GCA 158
Qy 398 ArgArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGln 417
Db 157 AGCGCGCTGTGGGGTGGCTGTGGAGGCTCAGATATCCGCAAGTGTGCTGCC- 104
Qy 418 ProLeuLysAlaArgGlyProSerAlaMetProAspProAlaGlnGly 433
Db 103 -----AGGGTGGATGGACCCCACTCACCCTGGTCCCGTGGGGC 62

```

RESULT 3

```

US-11-072-512-1688
; Sequence 1688, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1688
; LENGTH: 2624
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-072-512-1688
Alignment Scores:
Pred. No.: 3,43e-13 Length: 2624
Score: 302.00 Matches: 132
Percent Similarity: 37.7% Conservative: 76
Best Local Similarity: 24.0% Mismatches: 200
Query Match: 10.6% Indels: 144
DB: 11 Gaps: 23
US-10-635-977-2 (1-541) x US-11-072-512-1688 (1-2624)
Qy 82 ValGluAspLeuThrGluAlaGluTrpGlu-----91
Db 447 GTGGAGTGAAGGAGGAGGAGTGGATTCTACTGGTGTGACGTGAGTGGCTCCGG 506
Qy 92 -----AspLeuThr-----GlnGlnTyr 97
Db 507 GAGAACTTCGACCAACCTACATGGATGAACATGTGCGGATCAGTCACTTCGGAAACAC 566
Qy 98 TyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCys 117
Db 567 TATGAGTGAACCCGGAAGAACTACATGGTGAAGAACTTCCGAGCGGTTCGGAAGCAGCTG 626
Qy 118 GlnAlaLeuLeuAsnArgIleThrSerValAsn-----ProGlnThr-----131
Db 627 GAGCGTGAGCGAGGAAAGCTGGAGCGACCAAGTGTGACTTCTTCCCAAAACCTTTGAG 686
Qy 132 -----AspIleAspGlyLeuArgAsn-----IleTrp 140
Db 687 ATGCCTTTCGAGTACCACCTGTTTGTAGAGGAGTTTCGCAAAACCCAGGAATCACCTGG 746
Qy 141 IleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
Db 747 ATCATGAAGCCTGTAGCCCGCTCTCAAGGGAAAGGCATCTTCTCTTCCGTAGGCTGAAG 806
Qy 161 GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLys-----177
Db 807 GACATCTGGAGTGG-----AGGAAGGACACAGAAGAGCTCTGACGACCAAGATGATATT 863
Qy 178 -----TrpValValGlnLysTyrIleGluThrProLeuLeuLysCysAspThr 193
Db 864 CCCGTGGAGAACTATGTGGCTCAGCGTTACATCGAAATCCTTACCTGATAGGAGGCCGC 923
Qy 194 LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
Db 924 AAGTTTCACCTGCGTGTCTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 962
Qy 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuLeuAspLysLeuAspSer 233
Db 963 TTTGCTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1001
Qy 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253

```

```
Db      1002 GATGTTCACTCACCAACGTGGCTGTGCAAAA----- 1034
Qy      254 ProLeuLeuProAlaHisAsn-----MetTrpThrSerThrArgPheGln 268
Db      1035 ---ACATCTCCGAGTACTACCACCAAGAGGGCTGCAAGTGAGCGCTGCAGCGCTTCCGG 1091
Qy      269 GluTyLeu---GlnArgGlnGlyArgGlyAlaValTrpGlySerValIleTyProSer 287
Db      1092 CAGTACTCTCGGTCACCAACGCGGCCGAGGAGTG-----GAGACATCTTTCAGGGAC 1145
Qy      288 MetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAspHisValGluProArgLys 307
Db      1146 ATGCACAACATCTTTGTCAAAGCCTGCAGAGTGTGCAGAGGTGATCATCATGACACAAG 1205
Qy      308 AsnSerPheGluLeuTyArgAlaAspPheValLeuGlyArgAspPheArgProTrpLeu 327
Db      1206 CACTGCTTCGAGCTGTACGGCTATGACATCTCATCGACACGAGCACCTCAAGCGCTGGCTC 1265
Qy      328 IleGluLeuAsnSerSerProThrMetHisProSerThrProValThrAlaGlnLeuCys 347
Db      1266 CTGAGGTCATATCGTCCCATCACTGACAGCGCAGCAGCGAGGAGACTATGAGCTCAAG 1325
Qy      348 AlaGlnValGlnGluAspThrIleLysValAla-----ValAspArg 361
Db      1326 ACCTGCTCTCTGGAAGACACCTTGATGTTGTGGACATGGAACGAGGCTCAGCGGAAG 1385
Qy      362 SerCysAspIleGlyAsnGluLeuLeuTrpArgGlnProValValGlu----- 378
Db      1386 GAGAAGCGAGTGGGGCTTTTGACCTCATGTGAATGATGGCCCTGTAGCAGAGAGGAG 1445
Qy      379 -----ProProProPheSerGlySerAspLeuCysValAlaGlyValSerValArg 396
Db      1446 GGGGCTCTCTGA-CCTGTGCGGAATGGGAACTTTGTGACCAACACACATCTCGGTGCGT 1504
Qy      397 AlaArgArg----- 399
Db      1505 CAACGTCGGAAGAAACMACTGAGGAGCTCTTCTGCTCCTTCAAGTTTCAGAAAGC 1564
Qy      400 ---GlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnPro 418
Db      1565 TTCAGTGTGATCCCGAGCTGCCAGGAGGAATAATCAGCCTTAGC-----AGCGCT 1612
Qy      419 Leu-----LysAlaArgGlyProSerAlaMetProAspPro 430
Db      1613 TGTGTTTGTAGATTGGGAGCTGAGCCAGAAACCGGAAACCTCGGCTCATGGCGAGACA 1672
Qy      431 AlaGlnGlyProProSerProAlaLeuGlnArgAspLeuGlyLeuLysGluGly 450
Db      1673 GCTTCGGGAGTTGAGTGTCAAGCCCGAAGGGAAGACTTT---TGAAGGAAATACTGGC 1729
Qy      451 LeuProLeuAlaLeuLeuAlaProLeuArgGlyAlaAlaGlnSerGlyAlaAlaGln 470
Db      1730 AAATCCAAGAAGTGAACCGCGACTT---TAGGAAGCCAGAAGCCAGGCTACTCTAGA 1786
Qy      471 ProThrArgThrLysAlaAlaGlyLysValGluLeuProAlaCysProCysArgHisVal 490
Db      1787 CTTTCTCCCTTCAGAGCATGAGGACTTCTCCCACTATCCCTGTACTCAGGCGACACAA 1846
Qy      491 AspSerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTrpAspPro 510
Db      1847 CCCAGACAG-----GTCCAGGATGAGACTCCA 1873
Qy      511 AsnGlnLeuAsnAlaHisProLeuGluProValLeuArgGlyLeuLysThrAlaGluGly 530
Db      1874 GGTTCCTTA-----TTTCCAGAAGGAATGCGATTGCTTTTTTT 1912
Qy      531 AlaLeuArgProPro-----ProGlyGlyLysGly 540
Db      1913 GGCCAGACACCCCTGCCAGGGGGAAGGGG 1945
```

RESULT 4

US-11-072-512-1401

; Sequence 1401, Application US/11072512

```
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1401
; LENGTH: 2074
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-11-072-512-1401

Alignment Scores:
Pred. No.: 5,59e-11 Length: 2074
Score: 271.50 Matches: 107
Percent Similarity: 36.6% Conservative: 59
Best Local Similarity: 23.6% Mismatches: 149
Query Match: 9.5% Indels: 139
DB: 11 Gaps: 15

US-10-635-977-2 (1-541) x US-11-072-512-1401 (1-2074)
Qy      2 AlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerSerArgSer 21
Db      14 GCCTCCGCGCATTTTCAGCTGGCGCTGCAGC-----TGCTGCACAGAGACCCACAGAGC 67
Qy      22 LysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerSerArgGln----- 39
Db      68 CACCCTCGGAACACGAGCCCAATGAGAGGCGGGACCTGTGTCTCCACACAAAGCCAG 127
Qy      40 -----AspAla 41
Db      128 GCCTGGGATCTTTGAGAACCCACCACCCAGCCCTTACCCCTTAACATTCATCCGAGGCA 187
Qy      42 GluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLys 61
Db      188 AACCACTGAGCAGCGCCCTGCAGGCGCTGGGAGCAAGGCTACAGGAAGCAGGTGTTTCC 247
Qy      62 Val----- 62
Db      248 ATCCCTCCAGCGAGCGCCGCCAACACCAACTGGAGAGAGAGAAAAAACCTCATTTG 307
Qy      63 -----CysGlnAlaTyLeuGlyGlnLeuGluHisGluAspIleAsp 76
Db      308 ATGCGGAAAGATGAACTTCAGGGCCCTCTTGAAGCCGCTGGTTTTCGCGTTGACGAG 367
Qy      77 ThrSerAlaAspAlaValGluAsp-----LeuThrGluAlaGluTrpGluAspLeuThrGln 95
Db      368 ACCACCCCGGCTGTGGTGCAAAAGTGTCTCTCTGGAGAGGGGGTGGAAATAGTTTGTATAG 427
```

```
QY 96 GlnTyr-----TyrSerLeuValHisGlyAspAlaPheIleSerAsnSer 110
Db 428 CAGGAGCAGAACGCGGAGACTGGAACCTGTACTG-GAGGACATCCTCTTCGGAATGAC 486
QY 111 ArgAsnTyrPheSerGlnCys-----117
Db 487 CGACAC-----CAACAGTGTTAACCCGTGGCAGCAGCTAAACCACCCCTGGGAACAC 540
QY 118 GlnAlaLeuAsnArgIleThrSerValAsnPro-----129
Db 541 CAAGCTTACCAGGAAGACTGTTTGSCCAACACCTGAAGCACATCAGGAGGATGTATGG 600
QY 129 -----129
Db 601 CACTTCCCTGTACCAGTTTCATCCCCGTGACGTTGCTCATGCCCAATGACTATACCAAGTT 660
QY 130 -----GlnThrAsp-1leAspGlyLeuArgAsnIleTrpIleIrl 142
Db 661 CGTGGCTGNATCTTTCAGGAGAGGAGAGATGCTGGGCACCAAGCATAGCTATTGGATTG 720
QY 142 elysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGluIrl 162
Db 721 CAAGCTGCTGAGTTATCTCGTGGGAGGGGATACTAATTTTCAGTGCATTTAAAGACTT 780
QY 162 eLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValGlnLny 182
Db 781 CATC-----TTTGATGATATGTATACATAGTCAGAA 810
QY 182 sYrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPheIe 202
Db 811 ATATATCTCCAATCTTACTTATTGGCAGATATAAATGTGATCTCCGCATCTATGTTTG 870
QY 202 uValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPheSe 222
Db 871 TGTTACTGCTTAAAGCCTTTCACCATTTATGTTTATCAGGAAGGGTGGTTCGTTTGC 930
QY 222 rThrGlnArgPheSerLeuAspLysLeu---AspSerAlaIleHisLeuCysAsnAl 241
Db 931 CACGGAAGAGTTTGACCTCAGTAATTTGCMAACAATATATGCCCATTTGACCAACAGCAG 990
QY 241 aValGlnLys-----TyrLeuLysAsnAspValGlyArgSerProIe 255
Db 991 CATCAATAAATCCGGGGCTCTTATGAGAAGATCAAGAAGATGATGTGTCATGGT- 1045
QY 255 uLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnI 275
Db 1046 -----TGTAATGGACGCTCAGCAGATTTTTTTCCTACCTTCGTAGC----- 1087
QY 275 yArgGlyAlaValTrp-----GlySerValIleTyrProSerMetLysLysAlaIrl 292
Db 1088 -----TGGGATGTGGACGATCTGCTTTTGTGGAAGAAATCCACCGCATGGT 1134
QY 292 eAlaHisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluIe 312
Db 1135 TATTCTCACCATTCTCGCCATTGACCACTGTCCTCCCTTTGCTGCAATTCGTTTGAGCT 1194
QY 312 uTyrGlyAlaAspPheValLeuGlyArgPheArgProTrpLeuIleGluIleAsnSe 332
Db 1195 CTTTGGGTTGATATTTTGTATGATGATGACAACTTGAACCATGCTTTTGTAGAGTCACTA 1254
QY 332 rSerPro-----ThrMetHisProSerThrProVal 342
Db 1255 CAGCCACGCTTGACCTTGGATTGTTTCAACAGATGTG 1291
```

RESULT 5

```
US-11-072-512-1651
; Sequence 1651, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
```

```
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YUKI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1651
; LENGTH: 2111
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-072-512-1651
```

```
Alignment Scores:
Pred. No.: 2,6e-06 Length: 2111
Score: 210.00 Matches: 146
Percent Similarity: 32.7% Conservative: 53
Best Local Similarity: 24.0% Mismatches: 210
Query Match: 7.4% Indels: 210
DB: 11 Gaps: 31
US-10-635-977-2 (1-541) x US-11-072-512-1651 (1-2111)
```

```
QY 6 LeuLysTrpValValSerHisGlnSerCysSerArgSerArgSerLysProArgAsp 25
Db 532 CTGAAGTGG-----TGTCAGGTCAAGAGCCGACAGACGATACGCGCAGC 573
QY 26 GlnArgGluGluAlaGlySerSerAspLeuSerArgGlnAspAla----- 41
Db 574 TTCGGGAAGG-AGAGCAGCTGCTGTACCAGCTTCCCAACAACAAAGCTCCTCACCACCA 632
QY 42 GluAsnAlaGlu-AlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLy 61
Db 633 GATCGGGCTGCTCAGCACCTTTCGGGA---CGGCGACGGGCCATGACGAGGCCAGCA 689
QY 61 sValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAspAl 81
Db 690 GGTCCGGGGGGGTCCAGGCCAGGCTGGAAGAAG-----GACGCGAGCAGCGCCGC 740
QY 81 aValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrTyrSerLeuVa 101
Db 741 CCTGGAGACCTC-----CCGTGG-----759
QY 101 lHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeuLe 121
Db 760 -----ACAAGCCAGGATACCTCAGGCCACAGAGGGTCTGTG--- 795
QY 121 uAsnArgIleThrSerValAsnProGlnThr-----AspIleAspGlyLeuArg--- 137
Db 796 ---AGAATGAAGAGAGTTTTTCCAGAGACCTTACCGCTGGACCTCAAAACACGAGAGAGA 851
QY 138 -----AsnIleTrpIleIleLysProAlaAlaLysSe 148
Db 852 GGCCTTTTTCACCTGTTTGTATGAAGCCAGATATGATCTGCAAGCCCAACGAGCTCCA 911
```


[illegible]


```
Alignment Scores: 14.8 Length: 5354
Pred. No.: 126.00 Matches: 82
Score: 36.7% Conservative: 21
Percent Similarity: 36.7% Mismatches: 90
Best Local Similarity: 29.2% Indels: 88
Query Match: 4.4% Gaps: 20
DB: 14

US-10-635-977-2 (1-541) x US-11-136-527-1889 (1-5354)
QY 306 ArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArgAspPheArgPro 325
DB 4424 AGGAGGAATTCCTCAGAGTCCAGGAGGGGAC-----4456
QY 326 TrpLeuIleGluIleAsnSerProThrMetHisProSerThrProValThr-----343
DB 4457 -----CAAGCAACATCCCTTCACCTCCCTCGCGCTGCTT 4492
QY 344 ---AlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAspArgSer 362
DB 4493 CCCTCCAGCTCTCGCGGGCC-----ACCGTTCCAACTGCGTCTTCTTC 4543
QY 363 CysAspIleGlyAsnPheGluLeu-----TrpArgGlnProValValGluProPro 380
DB 4544 TGGGAGCAAGACTTTCAGATGTTGTAGCGTGGCGGCGACTCAGGCAGCGCCCTT 4603
QY 381 Pro-----PheSerGlySerAspLeuValAlaGlyValSerValArgArg 396
DB 4604 CCCCTAACGGCTCTTCTCCGCTCC-----ATTGCTCAGTTCCCTCCGCT 4651
QY 397 AlaArgArgGlnValLeuPro-ValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAl 416
DB 4652 -----CCTCCCTCTCTCT-----CCACCTCGGGAGCGGCTGGGGT 4690
QY 416 aGlnPro-----LeuLysAlaArgGlyPro-----SerAlaMe 427
DB 4691 GTGGCCAGGGGCGCGGTAAAGTCCCGGAGCGGTCCTCGGAGCGGCTCAGCCYCT 4750
QY 427 tProAspProAlaGlnGlyProProSerProAlaLeuGlnArgAspLeuGlyLeuLysG 447
DB 4751 GCCCCCGCGCTGCTGGCGCGCGGAGGATGACAGCGAGCGC---CTCGGTGTC---CA 4804
QY 447 uGluLysGlyLeuProLeuAlaLeuAlaProLeu-----ArgG 461
DB 4805 GGCTCGCTCGCTCAGCGTGTGCTAACTTCCCTTGGCTCGCTCCAGCTGCGGGG 4864
QY 461 yAlaAlaGluSerGly-----466
DB 4865 CCGCACCGTGTCTCTCGCTCTCGGGTCTGGTCTTAAGCTCTCTGGGAGTGCCTCC 4924
QY 467 -GlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysValGluLeuProAlaCys-- 485
DB 4925 GGGTCTCTTCAGCCCGCTCGGAGCGCGG-----CTCCCTTCTGCGC 4969
QY 486 ---ProCys-ArgHisValAspSerGlnAlaProAsnThrGlyValProValAlaGlnP 504
DB 4970 CCCACCTCGTCTCAGCATGCTCCCTTCGGCTCGT---GGCGCCCTGCTGCTGCC 5026
QY 504 roAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGluProValLeuArg 524
DB 5027 GCAGGGCTCGGCGAGCTGGGCA---CGAAGCATTCACCTGCCCTCTTGTCTCCGAG 5083
QY 524 lyLeuLysThrAlaGluGlyAlaLeuArgProProGly-----GlyLysGlySer 541
DB 5084 GA-----GAAGCTGGCGCGTGGCGCCCCCGGCTGTGTGAGGAGCTGGTGG 5133

RESULT 15
US-10-995-561-13381/c
; Sequence 13381, Application US/10995561
; Publication No. US2005027054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
```

```
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: C001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13381
; LENGTH: 14248
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13381

Alignment Scores: 38.5 Length: 14248
Pred. No.: 126.00 Matches: 49
Score: 34.1% Conservative: 13
Percent Similarity: 34.1% Mismatches: 64
Best Local Similarity: 26.9% Indels: 56
Query Match: 4.4% Gaps: 9
DB: 8

US-10-635-977-2 (1-541) x US-10-995-561-13381 (1-14248)
QY 375 ProValValGluProProPheSerGlySer-----AspLeuCysVal 389
DB 6063 CTTCCCGGAGCGCCCGCGCTTCGGATCCCCCTCCAGCTCTGCTCTCTGCTC 6004
QY 390 AlaGlyValSerValArgArgAlaArgGlnValLeuProValCysAsnLeuLysAla 409
DB 6003 CTTGGGATC-----5995
QY 410 SerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArgGlyProSerAlaMetProAsp 429
DB 5994 -----CTGGTTCCTGGAGTCACACCCACC 5968
QY 430 ProAlaGlnGlyProProSerProAlaLeuGlnArgAsp---LeuGlyLeuLysGluGlu 448
DB 5967 CCGGCTCAA---CCGGGCTCCCGGACCTCGCGGCGGCGGCTGACTCCGAGT 5911
QY 449 LysGlyLeuProLeuAlaLeuLeuAlaProLeuArgGlyAlaAlaGluSerGlyAla 468
DB 5910 CGCGGCTCCCGGCTCTGCTCGCGCGCGGCTCCGCTCCGCGCGGCTCC 5851
QY 469 AlaGlnProThrArgThrLysAlaAlaGlyLysValGluLeuProAlaCysProCysArg 488
DB 5850 GGCCCCCGCGCGCGCGCGC-----TGACCTGTCTCC 5818
QY 489 HisValAspSerGlnAlaProAsnThrGlyValProValAlaGlnPro---AlaLysSer 507
DB 5817 TGCTCTCAGCCAGCGCGCTCGCTCCCGCCCCCTCCCGCGCGCGCGCTCGCGGC 5758
QY 508 TrpAspProAsnGlnLeuAsnAlaHisProLeu-----GluProValLeu--- 522
DB 5757 TGGCGGCTCTCCGCTCTGCTCGCTCCGCTCGGTCGCGGAGACCGACCTGGGCTAGC 5698
QY 523 ArgGlyLeuLysThrAlaGluGlyAlaLeu-----ArgProProPro 536
DB 5697 AGGGGTCTAGACCGCGGAGGGGGCGGCACGAGGAGGAACCGGGCCCGGGACCGCCA 5638
QY 537 GlyGly 538
DB 5637 GGGGA 5632

Search completed: April 4, 2006, 12:29:52
Job time : 4850 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2006, 09:54:42 ; Search time 1192 Seconds
(without alignments)
3753.132 Million cell updates/sec

Title: US-10-635-977-2

Perfect score: 2854

Sequence: 1 MASSILKVVVSHQSCRSR.....LRGLKTARGALRPPPGKGS 541

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/abss/ABSSWEB_spool/US10635977/runat_04042006_103543_10259/app_query.fasta_1
-DB=Published Applications NA Main -QFMT=fastap -SUFFIX=p2n.rnpbm
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct
-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss07
-USER=US10635977 @CGN 1.1.1026 @runat_04042006_103543_10259 -NCPU=6 -ICPU=3
-NO MWAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA Main:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2854	100.0	1838	7	US-10-615-659-1
2	2854	100.0	1838	7	US-10-635-977-1
3	2843	99.6	3554	7	US-10-615-659-12
4	2843	99.6	3554	7	US-10-635-977-12
5	2698.5	94.6	1859	7	US-10-615-659-10
6	2698.5	94.6	1859	7	US-10-635-977-10
7	2693	94.4	3465	7	US-10-615-659-11

8	2693	94.4	3465	7	US-10-635-977-11	Sequence 11, Appl
9	2689	94.2	1939	7	US-10-615-659-9	Sequence 9, Appl
10	2689	94.2	1939	7	US-10-635-977-9	Sequence 9, Appl
c 11	1802	63.1	101270	8	US-10-723-860-631	Sequence 631, App
12	1299	45.5	726	7	US-10-615-659-3	Sequence 3, Appl
13	1299	45.5	726	7	US-10-635-977-3	Sequence 3, Appl
14	1070.5	37.5	2380	7	US-10-273-595A-39	Sequence 39, Appl
15	1020	35.7	2538	5	US-10-102-524-1765	Sequence 1765, Ap
16	1020	35.7	2553	9	US-10-956-149-1371	Sequence 2016, Ap
17	1020	35.7	2553	9	US-10-756-157-1371	Sequence 1971, Ap
18	1020	35.7	3001	5	US-10-037-270-494	Sequence 494, App
19	1020	35.7	3001	5	US-10-117-722-494	Sequence 494, App
20	1020	35.7	3001	9	US-10-122-851-494	Sequence 494, App
21	998	35.0	2611	3	US-09-925-298-199	Sequence 199, App
22	998	35.0	2611	3	US-10-102-806-199	Sequence 199, App
23	983.5	34.5	2848	6	US-10-108-260A-1766	Sequence 1766, Ap
24	926	32.4	1673	9	US-10-956-157-4121	Sequence 4121, Ap
25	912.5	32.0	1400	9	US-10-956-157-9356	Sequence 9356, Ap
26	785	27.5	1400	9	US-10-956-157-7251	Sequence 7251, Ap
c 27	760	26.6	490	6	US-10-029-386-24894	Sequence 24894, A
28	733	25.7	2979	10	US-11-097-143-23726	Sequence 23726, A
c 29	678	23.8	526	5	US-10-027-632-135928	Sequence 135928, A
c 30	678	23.8	526	6	US-10-027-632-135928	Sequence 135928, A
c 31	675	23.7	526	5	US-10-027-632-135927	Sequence 135927, A
c 32	675	23.7	526	6	US-10-027-632-135927	Sequence 135927, A
c 33	556.5	19.5	5728	10	US-11-097-143-23725	Sequence 23725, A
34	536.5	18.8	2543	10	US-11-097-143-23414	Sequence 23414, A
35	514.5	18.0	4615	10	US-11-097-143-23413	Sequence 23413, A
36	478.5	16.8	1958	9	US-10-956-157-4503	Sequence 4903, Ap
37	409.5	14.3	436	7	US-10-242-535A-7660	Sequence 7660, Ap
38	409.5	14.3	436	7	US-10-085-783A-7660	Sequence 7660, Ap
39	405	14.2	418	3	US-09-983-965-2032	Sequence 2032, Ap
40	394	13.8	1728	9	US-10-450-763-5962	Sequence 5962, Ap
41	390.5	13.7	2250	9	US-10-450-763-5187	Sequence 5187, Ap
42	364.5	12.8	755	5	US-10-027-632-135929	Sequence 135929, A
43	364.5	12.8	755	5	US-10-027-632-135930	Sequence 135930, A
44	364.5	12.8	755	6	US-10-027-632-135929	Sequence 135929, A
45	364.5	12.8	755	6	US-10-027-632-135930	Sequence 135930, A

ALIGNMENTS

RESULT 1

US-10-615-659-1
; Sequence 1, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (153)..(1775)
US-10-615-659-1

Alignment Scores:
Pred. No.: 9.2e-287 Length: 1838
Score: 2854.00 Matches: 541
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 7 Gaps: 0

Query Match:	100.0%	Indels:	0
DB:	7	Gaps:	0
US-10-635-977-2 (1-541) x US-10-635-977-1 (1-1838)			
QY	1	MetAlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerArg	20
DB	153	ATGGCATCAGCATCTCAAGTGGGTGTCACCCACAGAGCTGCAGCAGGAGCAGCA	212
QY	21	SerLysProArgAspGlnArgGluAlaGlySerSerAspLeuSerArgGlnAsp	40
DB	213	AGCAAGCCAGGAGCAGAGGAGGAGGCGGAGCAGCAGCTGAGCAGCAGGCAAGAT	272
QY	41	AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys	60
DB	273	GCTGAAATGCTGAGCAAGCTCAGGGGCTCCCGGGGACGCTTGTGACATCGCTGC	332
QY	61	LysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp	80
DB	333	AAGGTGTGCCAGCCCTACTTGGGAGCTGGAGCATGAGACATCGACATGTCAGCAT	392
QY	81	AlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrTrpSerLeu	100
DB	393	GCCGTGGAGACCTCACTGAGGCGGAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTC	452
QY	101	ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu	120
DB	453	GTTCAATGGCGATGCTTTCATCTCAATTCAGAAATTTACTTTTCGAGTGGCAGGCTCTG	512
QY	121	LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrp	140
DB	513	CTGAATAGATCACGTCTGTGAACCTCAGACGGACATGACGGGCTCCGGAACATCTGG	572
QY	141	IleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu	160
DB	573	ATTATAAGCCCGCGGCAAGTCCCGGGGCGGAGACATAGTGTGATGACCGTGTGGAG	632
QY	161	GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValVal	180
DB	633	GAGATCTTGGAGCTGGAGCTGCAGACCACTCTTTCAGGAGCAACAAGTGGGTGTC	692
QY	181	GlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrp	200
DB	693	CAGAAATACATCGAGCCCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGG	752
QY	201	PheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArg	220
DB	753	TTCTCTCGTCACGGACTGGAAACCCCTGACCATCTGGTTCTACAGGAGATTTACTTGGG	812
QY	221	PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn	240
DB	813	TTCTCACTCAGCGCTTCTCCTGGACAAAGCTGGACAGCGCCATCCACCTGTGCAACAAC	872
QY	241	AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn	260
DB	873	CGCGTCCAGAGATACCTGAAGAATGATGGGCGCAGCCCTGCTGCCCGCACACAC	932
QY	261	MetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTrp	280
DB	933	ATGTGGACACACAGGTTCCAGAGTACTTCGAGCGCCAGGCGCGTGGCGCGTGTGG	992
QY	281	GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln	300
DB	993	GGCAGCGTCATCTACCCCTCCATGAAGAAGGCCATCGCCACGCCATGAAGGTGGCCAG	1052
QY	301	AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly	320
DB	1053	GACCAGTGGAGCTCGCAAGACAGCTTTGAGTCTACGGGGCTGACTTCGCTCTTGGG	1112
QY	321	ArgAspPheArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThr	340
DB	1113	AGGACTTCAGGCCCTGGCTGATCGAGATCAATTCAGGCCCCACCATGACCCCTGCACG	1172
QY	341	ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp	360
DB	1173	CGGTCACGGCCAGCTGTGTGCACAGGTGCAGGAGGACACCATCAAGGTGCGCGTGGAC	1232
QY	361	ArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValValGluProPro	380
DB	1233	CGACGTGTGACATCGGCAACTTCGAGCTCTCTGTGTGAGGACGCGGTGTGAGCGCGCC	1292
QY	381	ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaArgArgGln	400
DB	1293	CCATTCAGCGGTCCGACCTCTGCGTGGGGCGGTCACTGTGTGAGGAGAGCCAGAGGAG	1352
QY	401	ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys	420
DB	1353	GTGCTGCCCGCTGTCAACCTCAAGGCTCGGCTCGCTGTTGGACGCGCAGCGCTGAAG	1412
QY	421	AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln	440
DB	1413	GCACGGGGCCCTCGGCCATGCCAGCCCTGCCAGGAGCCCTCATCACCAGCTCTCCAG	1472
QY	441	ArgAspLeuGlyLeuLysGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuArg	460
DB	1473	CGGACTTGGGACTGAAGAGAGAGAGGGGCTCCCTCGCTTGTGTGCACCCCTTAAG	1532
QY	461	GlyAlaAlaGluSerGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal	480
DB	1533	GGGGCAGCCGAGAGCGGTGGAGCCGACACAGCCGACCCCAAGCTGCTGGGAAGGTG	1592
QY	481	GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro	500
DB	1593	GAGCTCCCGGCTGCGCCCTGTGCCACGTGACAGTACGGGCCCAACACCGGTGTCCCC	1652
QY	501	ValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGluPro	520
DB	1653	GTAGCCAGCCCGCCAAAGCTGGATCCAAACCAAGCTAAATGCGCACCCGCTGGAGCT	1712
QY	521	ValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGlyLysGly	540
DB	1713	GTGCTGGGGGCTTGAAGACAGCAGAGGGCGCTGCTGCGCGCCCGGAGGAAAAGT	1772
QY	541	Ser 541	
DB	1773	TCA 1775	
RESULT 3			
US-10-615-659-12			
; Sequence 12, Application US/10615659			
; Publication No. US20040157234A1			
; GENERAL INFORMATION:			
; APPLICANT: Bristol-Myers Squibb Company			
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN			
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42			
; FILE REFERENCE: D0283 NP			
; CURRENT APPLICATION NUMBER: US/10/615,659			
; CURRENT FILING DATE: 2003-07-09			
; PRIOR APPLICATION NUMBER: U.S. 60/394,725			
; PRIOR FILING DATE: 2002-07-09			
; NUMBER OF SEQ ID NOS: 102			
; SOFTWARE: PatentIn version 3.2			
; SEQ ID NO 12			
; LENGTH: 3554			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-10-615-659-12			
Alignment Scores:			
Pred. No.:	3,24e-285	Length:	3554
Score:	2843.00	Matches:	539
Percent Similarity:	99.6%	Conservative:	0
Best Local Similarity:	99.6%	Mismatches:	2
Query Match:	99.6%	Indels:	0
DB:	7	Gaps:	0

US-10-635-977-2 (1-541) x US-10-615-659-12 (1-3554)	
QY	1 MetAlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerArg 20
Db	1695 ATGGCATCCAGCATCTCAAGTGGTGGTTCAGCCACAGAGCTGCAGCAGGACGACAGA 1754
QY	21 SerLysProArgAspGlnArgGluAlaGlySerSerAspLeuSerSerArgGlnAsp 40
Db	1755 AGCAACCCAGGAGCCAGAGGAGGAGGCGGAGCAGCAGCTGAGCAGCAGCGCAGAT 1814
QY	41 AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys 60
Db	1815 GCTGAAATGCTGAGGCAGCTCAGGGCCCTCCCGGGCAGCTTGTGGACATCGCGTGC 1874
QY	61 LysValCysGlnAlaTyrlleuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp 80
Db	1875 AAGGTGTCAGGCCCTACCTGGGGCAGCTGGAGCATGAGCAGCATCGACAGCAT 1934
QY	81 AlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrlSerLeu 100
Db	1935 GCGGTGGAGGACCTCACTGAGGCGGAGTGGAGGACCTGACCCAGCAGTACTACCTCC 1994
QY	101 ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrlPheSerGlnCysGlnAlaLeu 120
Db	1995 GTTCATGGCGATGCTTTCATCTCCAATTCAGAAATTAATTTTCGAGTGCCAGGCTCG 2054
QY	121 LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrp 140
Db	2055 CTGAATAGAAATCACTGTGAACCTTCAGACGACATTCAGCGGCTCCGGAAATCTCG 2114
QY	141 IleIleLysProAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
Db	2115 ATTATAAGCCCGCGCCCAAGTCCCGGGCCGAGACATAGTGTGCATGACCGCTGGAG 2174
QY	161 GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValVal 180
Db	2175 GAGATCTGGAGCTGGCAGCTGCAGACCACTCTTTCCAGGGCAACAAGTGGGTGTC 2234
QY	181 GlnLysTyrlleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrp 200
Db	2235 CAGAAGTACATCAGACGCGCGTCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 2294
QY	201 PheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrlLysGluSerTyrlLeuArg 220
Db	2295 TTCTCTGTCAGGACTGGAAACCCCTGACCATCTGGTTCACAAAGGAGATTACTTGGCG 2354
QY	221 PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn 240
Db	2355 TTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGACAGCGCCATCCACCTGTGCAACAC 2414
QY	241 AlaValGlnLysTyrlLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn 260
Db	2415 GCCGTCCAGAAGTACTCTGAAGAATGATGTGGCGCGCAGCCCTGTGTCGCCGACACAAC 2474
QY	261 MetTrpThrSerThrArgPheGlnGluTyrlLeuGlnArgGlnGlyArgGlyAlaValTrp 280
Db	2475 ATGTGGACCAACAGCTTCCAGGAGTACTTGACGGCCAGCGCCGCGGCGCGTGTGG 2534
QY	281 GlySerValIleTyrlProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln 300
Db	2535 GGACGGTCACTACCGCTCATGAAGAAGCCCATCGCCACCGCATGAAGGTGGCCACAG 2594
QY	301 AspHisValGluProArgLysAsnSerPheGluLeuTyrlGlyAlaAspPheValLeuGly 320
Db	2595 GACCACGTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGCTCTTGGG 2654
QY	321 ArgAspPheArgProTrpLeuIleGluLeuAsnSerSerProThrMetHisProSerThr 340
Db	2655 AGGAGCTTCAGGGCCCTGGCTGATCGAGATCAATTCAGCCGCCACCATGCGACCGCTCCAGC 2714
QY	341 ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp 360
Db	2715 CCGGTACGGCCCGAGCTGTGTGCACAGGTGCAGGAGGACCATCAAGGTGGCGGTGCAC 2774
US-10-635-977-2 (1-541) x US-10-635-977-12 (1-3554)	
QY	361 ArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValValGluProPro 380
Db	2775 CGCACTGTGACATCGCAACTTCAGAGCTCTCTGTGAGGACCGCGGTGGTTGAGCGCCC 2834
QY	381 PropheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaArgArgGln 400
Db	2835 CCATTCAGCGGTCCAGACCTCTGCGTGGCGGCGTCAGTGTGAGGAGACGAGGAGCAG 2894
QY	401 ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys 420
Db	2895 GTGCTGCCGCTGCAACCTCAAGGCTCGGCTCGCTGTGGACGCGCGCGCTGAAG 2954
QY	421 AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln 440
Db	2955 GCACGGGGCCCTCGGCCATGACACCTGCCCAGGGACCCCATCACAGCTCTCCAG 3014
QY	441 ArgAspLeuGlyLeuLysGluLysGlyLeuProLeuAlaLeuAlaProLeuArg 460
Db	3015 CGGGACTTGGGACTGAAGGAAGAGGGGCTCCCTCGCTTGTGCGCACCTTAAGG 3074
QY	461 GlyAlaAlaGluSerGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal 480
Db	3075 GGCGCAGCCGAGAGCGGTGGAGCCGACAGCCACCCGACCAAGCTCTGTGGAAAGGTG 3134
QY	481 GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro 500
Db	3135 GAGTCCCGGCTGCCCCCTGTGCACAGTGGACAGTACAGCCCCCAACACCGGTGTCCCC 3194
QY	501 ValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGluPro 520
Db	3195 GTAGCCACGCCCAAGCTGGGATCCAAACAGCTAAATGAGCACCCGCTGGAGCCT 3254
QY	521 ValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProProGlyLysGly 540
Db	3255 GTGCTCGGAGCTTGAAGACAGCAGAGGCGCGCTCGTCCGCCGCCGAGGAAAGGT 3314
QY	541 Ser 541
Db	3315 TCA 3317
RESULT 4	
US-10-635-977-12	
; Sequence 12, Application US/10635977	
; Publication No. US2004017131A1	
; GENERAL INFORMATION:	
; APPLICANT: Bristol-Myers Squibb Company	
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN	
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42	
; FILE REFERENCE: D0283A CIP	
; CURRENT APPLICATION NUMBER: US/10/635,977	
; CURRENT FILING DATE: 2003-08-07	
; PRIOR APPLICATION NUMBER: U.S. 60/394,725	
; PRIOR FILING DATE: 2002-07-09	
; PRIOR APPLICATION NUMBER: U.S.10/615,659	
; PRIOR FILING DATE: 2003-07-09	
; NUMBER OF SEQ ID NOS: 103	
; SOFTWARE: PatentIn version 3.2	
; SEQ ID NO 12	
; LENGTH: 3554	
; TYPE: DNA	
; ORGANISM: Homo sapiens	
US-10-635-977-12	
Alignment Scores:	
Pred. No.:	3,24e-285 Length: 3554
Score:	2843.00 Matches: 539
Percent Similarity:	99.6% Conservative: 0
Best Local Similarity:	99.6% Mismatches: 2
Query Match:	99.6% Indels: 0
DB:	Gaps: 0


```
Db 47 ATGGCATCCAGCATCTCAAGTGGGTGGTCAGCCACCAGAGCTGCAGCAGGACGACGAGA 106
Qy 21 SerLysProArgAspGlnArgGluAlaGlySerSerAspLeuSerSerArgGlnAsp 40
Db 107 AGCAAGCCAGGACACAGAGGAGGAGCGCGGAGCAGCGACCTGAGCAGCAGCGCAAGAT 166
Qy 41 AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspLeuAlaCys 60
Db 167 GCTGAATAATGCTCAGGCAAGCTCAGGGGGCTCCCGGGGAGGCTTGTGGACATCGCGTGC 226
Qy 61 LysValCysGlnAlaTrpLeuGlyGlnLeuGluHisGluAspLeuAspThrSerAlaAsp 80
Db 227 AAGGTGTGCAGAGCTTACCTGGGGCAGCTGGACATGAGGACATCGACAGCTCAGCAGAT 286
Qy 81 AlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTrpTyrSerLeu 100
Db 287 GCCGTGGAGACCTCACTGAGGCGGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC 346
Qy 101 ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu 120
Db 347 GTTCA-----TGCCAGGCTCTG 363
Qy 121 LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrp 140
Db 364 CTGATAGATACAGCTCTGTGAACCTCAGACGGACATTGACGGGCTCCGGAAACATCTGG 423
Qy 141 IleIleLysProAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
Db 424 ATTATAAGCCCGCGGCCAAGTCCCGGGCGGAGACATAGTGTGCATGGACCGGTGGAG 483
Qy 161 GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValVal 180
Db 484 GAGATCTGGAGCTGGCAGCTGCAGACCAACCTCTTTCCAGGGACCAACAAGTGGGTGTC 543
Qy 181 GlnLysTrpIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrp 200
Db 544 CAGAAGTACATCAGACGCGCTGCTCATCTGTGACACCAAGTTGACATCAGACAGTGG 603
Qy 201 PheLeuValThrAspTrpAsnProLeuThrIleTrpPheTrpLysGluSerTrpLeuArg 220
Db 604 TTCCTCGTCAGGACTGGAACCCCTGACCATCTGGTTCTACAAAGGAGAGTACTTGGCG 663
Qy 221 PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn 240
Db 664 TTCCTCACTCAGCGCTTCTCCCTGGACAGCTGACAGCGCCATCCACCTGTGCAACAC 723
Qy 241 AlaValGlnLysTrpLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn 260
Db 724 GCCGTCCAGAAGTACCTGAAGATGATGTGGCGCGCAGCCCTGCTGCTGCCGCGACACA 783
Qy 261 MetTrpThrSerThrArgPheGlnGluTrpLeuGlnArgGlnGlyArgGlyAlaValTrp 280
Db 784 ATGTGGACCAAGCAGCAGGTTCCAGAGTACTCTGCAGCGCCAGGGCGCTGGCGCTGG 843
Qy 281 GlySerValIleTrpProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln 300
Db 844 GGCAGCGCTCATCTACCGCTCATGAAGAGCCATCGCCACGCGCAAGAGGTGGCCCGAG 903
Qy 301 AspHisValGluProArgLysAsnSerPheGluLeuTrpGlyAlaAspPheValLeuGly 320
Db 904 GACCACGTGGAGCCTCCGAAGACAGCTTTGAGGCTCTACGGGGCTGACTTGCCTTGGG 963
Qy 321 ArgAspPheArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThr 340
Db 964 AGGGACTTACGGCCCTGGCTGATCGAGATCAATTCAGCGCCACCATGACCCCGTCCAGC 1023
Qy 341 ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp 360
Db 1024 CCGGTACAGGCCCGAGCTGTGTGCAAGTGCAGGAGGACACCATCAAGGTGGCGCTGGAC 1083
Qy 361 ArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValValGluProPro 380
Db 1084 CGCAGCTGTGACATCGGCNACTTCGAGCTCTCTGTGGAG-----CCGGTGGTTGACCCGCC 1139
```

```
Qy 381 ProPheSerGlySerAspLeuCysValAlaIleGlyValSerValArgArgAlaArgArgGln 400
Db 1140 CCATTTCAGCGGGTCCGACCTCTGCGTGGCGGGCGTCAGTGTGAGGAGACCGAGGAGCAG 1199
Qy 401 ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys 420
Db 1200 GTGCTGCCCGTCTGCAACCTCAAGGCTCGGCTCGCTGTTGGACGCGCAGCGCTGAAG 1259
Qy 421 AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln 440
Db 1260 GCACGGGGCCCTCGGCCATGCGACACCTGCGCCAGGGACCCCATCACCAGCTCTCCAG 1319
Qy 441 ArgAspLeuGlyLeuLysGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuArg 460
Db 1320 CGGGACTTGGGACTGAAGGAAGAGAGGGGCTCCCTCGCTTGTGTGGACCCCTTAAGG 1379
Qy 461 GlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal 480
Db 1380 GGGGACCGCGAGAGCGGTGGACCGCACAGCCCGCACCAAGCTGCTGGGAAGGTG 1439
Qy 481 GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro 500
Db 1440 GAGCTCCCGGCTGCGCCCTGTCGCCACGTGGACAGTCAGGCCCAACACCGGTGCCCC 1499
Qy 501 ValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGluPro 520
Db 1500 GTAGCCCGCGCCCAAGCTGGGATCCAAACAGTCAGGCCCAACACCGCTGGAGCCT 1559
Qy 521 ValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGlyLysGly 540
Db 1560 GTGCTCGGAGCTGAAGACAGAGGGCGCGCTGCTCGCGCCCGCGAGGAAAGGT 1619
Qy 541 Ser 541
Db 1620 TCA 1622
```

RESULT 6

```
US-10-635-977-10
; Sequence 10, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-10
```

```
Alignment Scores:
Pred. No.: 1,52e-270 Length: 1859
Score: 2698.50 Matches: 522
Percent Similarity: 96.5% Conservative: 0
Best Local Similarity: 96.5% Mismatches: 4
Query Match: 94.6% Indels: 17
DB: 2
```

US-10-635-977-2 (1-541) x US-10-635-977-10 (1-1859)

```
Qy 1 MetAlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerArg 20
Db 47 ATGGCATCCAGCATCTCAAGTGGGTGGTCAGCCACCAGAGCTGCAGCAGGAGCAGAGA 106
```


Db	1738	AGCAAGCCACAGGACACAGAGGAGGAGGCCGGAGCAGCAGCTGAGCAGCAGGCAAGAT	1797	QY	401	ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys	420
QY	41	AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys	60	Db	2811	GTGTGCCCCGTCTGCAACCTCAAGGCCCTCGGCTCGTGTGGACCGCGCGTGAAG	2870
Db	1798	GCTGAAATGCTGAGCAAGCTCAGGGGCTCCCGGGCAGCTTGTGGACATCGCGTGC	1857	QY	421	AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln	440
QY	61	LysValCysGluAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp	80	Db	2871	GCACGGGGCCCTCGGCCATGCCAGCCTGCCAGGAGCCCCCATCACAGCTCTCCAG	2930
Db	1858	AAAGTGTGTCAGGCTTACCTGGGGCAGCTGGAGCATGAGGACATGACACGTGACAGAT	1917	QY	441	ArgAspLeuGlyLeuLysGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuArg	460
QY	81	AlaValGluAspLeuThrGluAlaGluTyrGluAspLeuThrGlnGlnTyrTyrSerLeu	100	Db	2931	CGGGACTTGGGACTCAAGGAGAGAGGGGCTCCCTCGCTTGTGTCACCTTAAGG	2990
Db	1918	GCGTGGAGGACTTACTGAGCCGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC	1977	QY	461	GlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal	480
QY	101	ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu	120	Db	2991	GGGGCAGCCGAGAGCGGTGGAGCCGACAGCCACCCGACCAAGCTCTGGGAAGGTG	3050
Db	1978	GTTTCATGGCGATGCTTTCATCTCCATTCCAATTCAAGAAATTACTTTTCGAGTGCCAGGCTCG	2037	QY	481	GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro	500
QY	121	LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTyr	140	Db	3051	GAGCTCCCGCCCTGCCCCCTGTGCCACGTGGACAGTCAGGCCCAACACCGGTGTCCCC	3110
Db	2038	CTGAATAGAATCACGTCTGTGAACCTCAGACCGCAATTCGCGGGCTCCGGAACATCTGG	2097	QY	501	ValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGluPro	520
QY	141	IleIleLysProAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu	160	Db	3111	GTAGCCAGCCGCCCAAAAGCTGGGATCCAAACCAAGCTAAATGAGCACCCGCTGGAGCCT	3170
Db	2098	ATTATAAGCCCGCGCCCAAGTCCCGGGCGG-----	2129	QY	521	ValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGlyLysGly	540
QY	161	GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValVal	180	Db	3171	GTGCTGCGGAGCTTGAAGACAGCAGAGGGCGCGCTCGTCCGCCCGGAGGAAAAGGT	3230
Db	2130	-----AGGGACAACAAGTGGGTGGTC	2150	QY	541	Ser	541
QY	181	GlnLysTyrIleGluThrProLeuLeuIleCysAspThrIlyAspIleArgGlnTyr	200	Db	3231	TCA	3233
Db	2151	CAGAAGTACATCAGACGCGCGTGTCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG	2210	RESULT 8			
QY	201	PheLeuValThrAspTrpAsnProLeuThrIleTyrPheTyrLysGluSerTyrLeuArg	220	US-10-635-977-11			
Db	2211	TTCCCTCGTCAGGACTGGAAACCCCTGACCACTGGTTCTACAAAGAGAGTACTTTCGGG	2270	; Sequence 11, Application US/10635977			
QY	221	PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn	240	; Publication No. US20040171131A1			
Db	2271	TTTCTCAACTCAGCGCTTCTCCCTGGACAGCTGACAGCGCCATCCACCTGTGCCAACAC	2330	; GENERAL INFORMATION:			
QY	241	AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn	260	; APPLICANT: Bristol-Myers Squibb Company			
Db	2331	GCCGTCAGAAAGTACCTGAAGAAATGATGTGGCGCGCAGCCCTGCTGCCCGCACACAAC	2390	; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN			
QY	261	MetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTyr	280	; FILE REFERENCE: D0283A CIP			
Db	2391	ATGTGGACCAAGCAGGTTCCAGGAGTACTTGAGCGCCAGGCGCGTGGCGCGTGGTG	2450	; CURRENT APPLICATION NUMBER: US/10/635,977			
QY	281	GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln	300	; PRIOR FILING DATE: 2003-08-07			
Db	2451	GGCAGCGTCATCTACCGCTCCATGAAGAAGCCATCGCCACCGCATGAAGTGGCCACAG	2510	; PRIOR FILING DATE: 2002-07-09			
QY	301	AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly	320	; PRIOR FILING DATE: 2003-07-09			
Db	2511	GACCACGTGAGCGCTCGCAAGAACAGCTTTGAGCTCTACGCGGCTGACTTCGCTTGGG	2570	; NUMBER OF SEQ ID NOS: 103			
QY	321	ArgAspPheArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThr	340	; SOFTWARE: PatentIn version 3.2			
Db	2571	AGGAGCTTCAGGGCCCTGGCTGATCGAGATCAATTCAGCCCCCATGCAACCGTCCACG	2630	; SEQ ID NO 11			
QY	341	ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp	360	; LENGTH: 3465			
Db	2631	CCGGTCACGGCCACAGCTGTGTGCACAGGTGCAGGAGGACCATCAAGTGGCGCTGGAC	2690	; TYPE: DNA			
QY	361	ArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValValGluProPro	380	; ORGANISM: Homo sapiens			
Db	2691	CGCAGCTGTGACATCGGCAACTTCGAGCTCTGTTGGAGGCGCGGTGGTGTAGCCGCC	2750	US-10-635-977-11			
QY	381	ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaArgArgGln	400	Alignment Scores:			
Db	2751	CCATTACGGGGTCCGACCTCTGCTGGCGGGCGTCAAGTGTGAGGAGAGCCAGGCGAG	2810	Pred. No.: 1.36e-269 Length: 3465			
				Score: 2693.00 Matches: 516			
				Percent Similarity: 95.4% Conservative: 0			
				Best Local Similarity: 95.4% Mismatches: 3			
				Query Match: 94.4% Indels: 23			
				DB: 7 Gaps: 1			
				US-10-635-977-2 (1-541) x US-10-635-977-11 (1-3465)			
				QY 1 MetAlaSerSerIleLeuLysTyrValValSerHisGlnSerCysSerArgSerArg 20			
				Db 1678 ATGCGATCCAGCATCTCTCAAGTGGTGTGAGCCACAGAGCTGCAGCAGGAGCAGCAGA 1737			
				QY 21 SerLysProArgAspGlnArgGluAlaGlySerSerAspLeuSerSerArgGlnAsp 40			
				Db 1738 AGCAAGCCAGGAGCAGAGAGGAGGCGCGGAGCAGCAGCTGTGAGCAGGAGCAAGAT 1797			

```
QY 41 AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys 60
Db 1798 GCTGAAATATGCTCAGGCAAAAGCTCAGGGGCTCCCGGGCAGCTTGTGGACATCGCGTGC 1857
QY 61 LysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp 80
Db 1858 AAGGTGTGCCAGCCCTACCTTGGGGCAGCTGGAGCATGAGACATCCACACGTCAGCAGAT 1917
QY 81 AlaValGluAspLeuThrGluAlaGluTyrGluAspLeuThrGlnGlnTyrIleValSerLeu 100
Db 1918 GCCGTGGAGACCTCACTGAGGCCGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC 1977
QY 101 ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu 120
Db 1978 GTTCATGGGAGTCTTTCATCTCCAATTCAAGAAATTAATTTTCGCGAGTGCAGGCTCTG 2037
QY 121 LeuAsnArgIleThrSerValLeuProGlnThrAspIleAspGlyLeuArgAsnIleTyr 140
Db 2038 CTGAATAGAAATCACGTCGTGTGAACCTCTCAGACGGACATTTGACGGGCTCCGGAACATCTGG 2097
QY 141 IleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
Db 2098 ATTATTAAGCCCGGGCCAGTCCCGGGCGG----- 2129
QY 161 GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTyrValVal 180
Db 2130 -----AGGACAAACAAGTGGTGGTC 2150
QY 181 GlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTyr 200
Db 2151 CAGAAATACATCAGACGCGCTGCTCATCTGTGACACCAAGTTCCGACATCAGACAGTGG 2210
QY 201 PheLeuValThrAspTyrAsnProLeuThrIleTyrPheTyrLysGluSerTyrLeuArg 220
Db 2211 TTCCTCGTCAAGACTGGAAACCCCTGACCATCTGGTTCACAGGAGAGTTACTTGGCG 2270
QY 221 PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn 240
Db 2271 TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGTCAACAAC 2330
QY 241 AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn 260
Db 2331 GCCGTGCCAAGATCTCTGAAGAATGATGTGGCCCGCAGCGCCCTGTGTCGCCGCACACAAC 2390
QY 261 MetTyrThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTyr 280
Db 2391 ATGTGACCCAGCACCAAGTTCCAGGAGTACCTGCAGCGCCAGCGCGCTGGCGCGGTGGTGG 2450
QY 281 GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln 300
Db 2451 GGCAGCGTCATCTACCGCTCCATGAAGAAGGCCATCGCCACGCCATGAAGTGGCCCGAG 2510
QY 301 AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly 320
Db 2511 GACCACTGGAGCCCTCGAAGAACAAGCTTTGAGCTCTACGGGCTGACTTCGTCCTCTGGG 2570
QY 321 ArgAspPheArgProTyrLeuIleGluIleAsnSerSerProThrMetHisProSerThr 340
Db 2571 AGGGACTTCAGGCCCTCGCTGTGATCAGATCAATTCAGGCCCCCATGTCACCCGCTCCAG 2630
QY 341 ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp 360
Db 2631 CCGGTCCAGGCCCGAGCTGTGTGCACAGTGCAGGAGGACCATCAAGGTGGCCGTGGAC 2690
QY 361 ArgSerCysAspIleGlyAsnPheGluLeuLeuTyrArgGlnProValValGluProPro 380
Db 2691 CGCAGCTGTGACATCGCAACTCTCAGACTCTCTGTGGAGCGACCGCGTGTGGCGGCC 2750
QY 381 ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgArgAlaArgGln 400
Db 2751 CCATTCAGCGGTCTCGACCTCTGCGTGGCGGCGCTGAGTGTGAGGAGCCAGGAGGCGAG 2810
```

```
QY 401 ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys 420
Db 2811 GTGCTGCCGCTCTGCAACCTCAAGGCTCGGCTCGCTGTTGACGCGCAGCGCTGAAG 2870
QY 421 AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln 440
Db 2871 GCACGGGGCCCTCGGCCATGCGACACCTGCCAGGACCCCCATCACCAGCTCTCCAG 2930
QY 441 ArgAspLeuGlyLeuLysGluGlyLysGlyLeuProLeuAlaLeuLeuAlaProLeuArg 460
Db 2931 CGGGACTTGGGACTGNAAGGAGAGAGGGGCTCCCTCGCTGGCTTGTGGCACCCTTAAGG 2990
QY 461 GlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal 480
Db 2991 GGGGAGCGCAGAGCGGTGAGCGCGCACAGCCACCGCAAAAGCTGCTGGGAAGGTG 3050
QY 481 GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro 500
Db 3051 GAGCTCCCGGCTCGCCCTGTGCGCACGTGCGGAGTCAAGTCAAGTCAAGTCAAGTCAAGT 3110
QY 501 ValAlaGlnProAlaLysSerTyrAspProAsnGlnLeuAsnAlaHisProLeuGluPro 520
Db 3111 GTAGCCAGCCCGCCCAAAAGCTGGATCCAAACCACTAAATGAGCACCCGCTGGAGCCT 3170
QY 521 ValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGlyLysGly 540
Db 3171 GTGCTCGGAGCTTGAAGACAGACAGAGGGCGGCTGCGTCCGCGCCCGGAGGAGGAGT 3230
QY 541 Ser 541
Db 3231 TCA 3233
```

RESULT 9

```
US-10-615-659-9
; Sequence 9, Application US/10615659
; Publication NO. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-9
```

Alignment Scores:

Pred. No.:	1.58e-269	Length:	1939
Score:	2689.00	Matches:	515
Percent Similarity:	95.2%	Conservative:	0
Best Local Similarity:	95.2%	Mismatches:	4
Query Match:	94.2%	Indels:	23
DB:	7	Gaps:	1

US-10-635-977-2 (1-541) x US-10-615-659-9 (1-1939)

```
QY 1 MetAlaSerSerIleLeuLysTyrValValSerHisGlnSerCysSerArgSerArg 20
Db 153 ATGGCATCCAGCATCTCAAGTGGGTGTCTAGCCACCCAGAGCTGCAGCAGGAGCAGAGA 212
QY 21 SerLysProArgAspGlnArgGluAlaGlySerSerAspLeuSerSerArgGlnAsp 40
Db 213 AGCAAGCCAGGGACCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 272
QY 41 AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys 60
```



```
Db      273  GCTGAAATGCTGAGGCAAGCTCAGGGGCTCCCGGGCAGCTTGTGGACATCGCGTGC 332
Qy      61   LysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp 80
Db      333  AAGGTGTGCGCAGCCCTACCTGGGCGAGCTGGAGCATGAGACATCGACAGTTCAGCAGAT 392
Qy      81   AlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrTyrSerLeu 100
Db      393  GCCGTGGAGACCTCACTGAGGCCGAGTGGGAGACCTGAGCCAGAGTAGTACTACTCCCTC 452
Qy     101   ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu 120
Db     453  GTTCATGGCGATGCTTCACTCTCAATTCAAGAAATTACTTTTCGAGTGCAGGCTCTG 512
Qy     121   LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrp 140
Db     513  CTGAATAGAAATCAGCTCTGTGAACCTCTCAGACGGACATTGACGGGTCCCGAACTCTGG 572
Qy     141   IleIleIysProAlaAlaIysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
Db     573  ATTATAAGCCCGCGCAAGTCCCGGGCGC----- 604
Qy     161   GluIleLeuGluAlaAlaAlaAspHisProLeuSerArgAspAsnLysTrpValVal 180
Db     605  -----AGGACAAACAAGTGGTGGTC 625
Qy     181   GlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrp 200
Db     626  CAGAAGTACATCAGACGCCGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 685
Qy     201   PheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArg 220
Db     686  TTCTCTGTCGCGAGTGAACCCCTGACCATCTGGTTCTACAGGAGAGATTACTTGGCG 745
Qy     221   PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn 240
Db     746  TTCTCAACTCAGCGCTTCTCCCTGGCAAGCTGGACAGCGCCATCCACCTGTGCAACAC 805
Qy     241   AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn 260
Db     806  ACCGTCCAGAAGTACCTGAAGAATGATGTGGCGCGCAGCCCTCTGCTGCCGCACACAAC 865
Qy     261   MetTrpThrSerThrArgPheGlnTyrLeuGlnArgGlnGlyArgGlyAlaValTrp 280
Db     866  ATGTGGACACGACACAGGTTCCAGGAGTACTCTGCAGCGCCAGGGCGCTGGCGCGTGG 925
Qy     281   GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln 300
Db     926  GGCAGCGTCACTACCCGTCATGAAGAAGGCCATCGCCACGCCATGAAGTGGCCCGAG 985
Qy     301   AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly 320
Db     986  GACCACGTGGAGCCTCGCAAGACAGCTTTGAGCTCTACGGGGTGACTTCGTCTTGGG 1045
Qy     321   ArgAspPheArgProTrpLeuIleGluLeuAsnSerSerProThrMetHisProSerThr 340
Db    1046  AGGAGACTTCAGGGCTCGGCTGATCGAGATCAATTTCCAGCCGCCACCATGACCCCGTCCAG 1105
Qy     341   ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp 360
Db    1106  CCGGTACGGCCCGAGCTGTGTGCACAGTGCAGAGGACACCATCAAGTGGCGCTGGAC 1165
Qy     361   ArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValValGluProPro 380
Db    1166  CGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGGAGGCAGCGGTGGTTGAGCCGCC 1225
Qy     381   ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaArgArgGln 400
Db    1226  CCATTACGGGGTCCGACCTCTGGCTGGCGGGCGTCAGTGTGAGGAGAGCCAGAGGCAG 1285
Qy     401   ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys 420
Db    1286  GTGCTGCCCGTCTGCAACCTCAAGGCCCTCGGCCCTCGCTGTGTGGACGGCGCCGCTGAG 1345

; Sequence 9, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BG942
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-9

Alignment Scores:
Pred. No.: 1.58e-269 Length: 1939
Score: 2689.00 Matches: 515
Percent Similarity: 95.2% Conservative: 0
Best Local Similarity: 95.2% Mismatches: 4
Query Match: 94.2% Indels: 23
DB: 7 Gaps: 1

US-10-635-977-2 (1-541) x US-10-635-977-9 (1-1939)
Qy      1   MetAlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerArg 20
Db     153  ATGCGATCCAGCATCTCTCAAGTGGTGGTTCAGCCACCAGAGCTCCAGCAGGAGCAGAGA 212
Qy     21   SerLysProArgAspGlnArgGluAlaGlySerSerAspLeuSerSerArgGlnAsp 40
Db     213  AGCAAGCCCGAGGACCAAGAGAGAGAGCGCGGAGAGCGGACCTGTGAGCAGCAGCAAGAT 272
Qy     41   AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys 60
Db     273  GCTGAAATGCTGAGGCAAGCTCAGGGGCTCCCGGGCAGCTTGTGGACATCGCGTGC 332
```


QY 61 LysValCysGlnAlaTyrLeuGlnLeuGluHisGluAspThrSerAlaAsp 80
 DB 333 AAGGTGTGCAGCCCTACCTGGGGCAGCTGGAGCATGAGCATGCACACGT CAGCAGAT 392
 QY 81 AlaValGluAspLeuThrGluAlaGluTyrGluAspLeuThrGlnGlnTyrTyrSerLeu 100
 DB 393 GCCGTGGAGACCTTACTAGGCCGAGTGGAGGACCTGACCAGCAGTACTTCCCTC 452
 QY 101 ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu 120
 DB 453 GTTCATGGCGATGCTTTCATCTCCAATTCAGAAATTAATTTTCGAGTGCAGGCTCTG 512
 QY 121 LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTyr 140
 DB 513 CTGAATAGAAATCACGTCTGTGAACCTTCAGACGGAATTCAGCGGCTCCGGAACATCTCG 572
 QY 141 IleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
 DB 573 ATTATTAAGCCCGCGCCGANGTCCCGGGCCG----- 604
 QY 161 GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTyrValVal 180
 DB 605 -----AGGACACACAGTGGGTGTC 625
 QY 181 GlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTyr 200
 DB 626 CAGAAATGATCATCGACGCGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 685
 QY 201 PheLeuValThrAspTyrAsnProLeuThrIleTyrPheTyrLysGluSerTyrLeuArg 220
 DB 686 TTCCTCGTCAGGAGCTGAACCCCTGACCATCTGGTTCTACAGGAGAGTTACTTGGGG 745
 QY 221 PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn 240
 DB 746 TTCTCAATTCAGCGTTTCTCCTGACAGCTGGACGCGCATCCACTGTGCAACAC 805
 QY 241 AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn 260
 DB 806 ACCGTCCAGAAAGTACCTGAAGAATGATGTGGCCCGCAGCCCTCTCTCCCGCACACAC 865
 QY 261 MetTyrThrSerThrArgPheGlnGlnTyrLeuGlnArgGlnGlyArgGlyAlaValTyr 280
 DB 866 ATGTGACACGACACAGGTTCAGGAGTACCTGTCAGCGCCAGGCGCGTGGCGCGCTGGG 925
 QY 281 GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln 300
 DB 926 GGCAGCGTCATCTACCCGTCCATGAGAGGCGCATCGCCCGCATGAAGTGGCCCGAG 985
 QY 301 AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly 320
 DB 986 GACCAGTGGAGCTTCGCAAGAACAGCTTTGAGCTCTACGGGCTGACTTCGTCTCTGGG 1045
 QY 321 ArgAspPheArgProTyrPheLeuIleGluLeuAsnSerSerProThrMetHisProSerThr 340
 DB 1046 AGGGAATTCAGGCGCTGCTGATCGAGATCAATTCAGCCCGCCACCATGACCCGCTCCACG 1105
 QY 341 ProValThrAlaGlnLeuCysAlaGlnValGlnAspThrIleLysValAlaValAsp 360
 DB 1106 CCGTTCAGCGCCAGCTGTGTGACAGGTGACGAGGAGACCATCAGGTGGCGGTGGAC 1165
 QY 361 ArgSerCysAspIleGlyAsnPheGluLeuLeuTyrArgGlnProValValGluProPro 380
 DB 1166 CGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGGAGCGAGCGGTGGTGGAGCGGCC 1225
 QY 381 ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaArgGln 400
 DB 1226 CCATTCAGCGGTTCGACCTCTGGCTGGCGGCGTCACTGTGAGAGAGCCAGGAGGCGAG 1285
 QY 401 ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys 420
 DB 1286 GTGCTGCCCGTGTGCAACTCAAGGCGCTCGGCGCTGCTGTGTGGAGCGCGCAGCCGCTGAAG 1345

QY 421 AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln 440
 DB 1346 GCACGGGGCCCCCTCGCCATGCGAGACCCCTGCCAGGAGACCCCATCACCAGCTCTCCAG 1405
 QY 441 ArgAspLeuGlyLeuLysGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuArg 460
 DB 1406 CGGGACTTGGGACTGAAGGAAGAGAGGGGCTCCCGCTTGTGTCGACCCCTTAAGG 1465
 QY 461 GlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal 480
 DB 1466 GGGGACGCGAGAGCGGTGGAGCCGCACAGCCACCCGACCAAGCTGCTGGGAAGGTG 1525
 QY 481 GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro 500
 DB 1526 GAGCTCCCGGCGCTGCCCTGTGCCACGCTGGACAGTCAAGCCCAACACCCGCTGCCCC 1585
 QY 501 ValAlaGlnProAlaLysSerTyrAspProAsnGlnLeuAsnAlaHisProLeuGluPro 520
 DB 1586 GTAGCCCAAGCCCGCCAAAGCTGGGATCCAAACACAGCTAAATGAGCACCCGCTGGAGCCT 1645
 QY 521 ValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGlyLysGly 540
 DB 1646 GTGCTGCGGAGCCTGAACAGACAGAGGGCGGCTGCGTCCGCCCGCGAGGAGAAAGT 1705
 QY 541 Ser 541
 DB 1706 TCA 1708

RESULT 11

US-10-723-860-631/c
 ; Sequence 631, Application US/10723860
 ; Publication No. US20040253606A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Aziz, Nataasha
 ; APPLICANT: Ginsburg, Wendy M.
 ; APPLICANT: Zlotnik, Albert
 ; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
 ; FILE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
 ; FILE REFERENCE: 05882.0193.NPUS01
 ; CURRENT APPLICATION NUMBER: US/10/723,860
 ; CURRENT FILING DATE: 2003-11-26
 ; PRIOR APPLICATION NUMBER: 60/429,739
 ; PRIOR FILING DATE: 2002-11-26
 ; NUMBER OF SEQ ID NOS: 8393
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 631
 ; LENGTH: 101270
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-723-860-631

Alignment Scores:
 Pred. No.: 1,18e-174 Length: 101270
 Score: 1802.00 Matches: 438
 Percent Similarity: 32.2% Conservative: 1
 Best Local Similarity: 32.1% Mismatches: 3
 Query Match: 63.1% Indels: 923
 DB: 8 Gaps: 3

US-10-635-977-2 (1-541) x US-10-723-860-631 (1-101270)

QY 100 LeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAla 119
 DB 47222 CTTTGTAGTGGGATGCTTTCATCTCCAAATTCAGAAATTAATTTTCGACGTCCAGGCT 47163
 QY 120 LeuLeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIle 139
 DB 47162 CTGCTGAATAGAAATCACGTCTGTGAACCCCTCAGACGACATTGACGGGCTCCGGAACATC 47103
 QY 140 TrpIleIleLysProAlaAlaLysSerArgGlyArg----- 151
 DB 47102 TGGATTATAAAGCCCGCGCCAGTCCGGGGCCG-AGGTGAGTCCCTGCTGGCGATGC 47044


```
Db 44883 GACCGSCCTTGGTGCCCTCCTGGGCCAAGGAGGCTGAGCACCTGCCCTCGTCCCCC 44824
QY 233 ---SerAlaIleHisLeuCysAsnAenAlaValGlnLysTyLeuLysAsnAspValGly 251
Db 44823 TGCAGCGCCATCCACCTGTGCAACAACGCGCTCCAGAGTACTGAGAAATGATGTGGC 44764
QY 252 ArgSerProLeuLeuProAlaHisAsnMetTTPThrSerThrArgPheGlnGluTyLeu 271
Db 44763 CCGAGCCCTCCTGCTGCCGCCACACAATGTGGACGAGCACCAGGTTCAGAGATCCCTG 44704
QY 272 GlnArgGlnGlyArgGlyAlaValTTPGlySerValIleTyProSerMetLysLysAla 291
Db 44703 CAGCGCCAGGCGCGTGGCCGCTGTGGGCGAGCTCATCTACCGCTCCATGAAGAGGCC 44644
QY 292 IleAlaHisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGlu 311
Db 44643 ATCGGCCACGCCATGAAGGTGGCCAGGACACGCTGGAGCCTCGCAAGAACAGCTTTGAG 44584
QY 312 LeuTyGlyAlaAspPheValLeuGlyArgAspPheArgProTTPLeuIleGluIleAsn 331
Db 44583 CTCTACGGGGCTGACTTCTGCTTGGAGGGGACTTTCAGGCCCTTGCTGATCGAGATCAAT 44524
QY 332 SerSerProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGln 351
Db 44523 TCCAGCCCCACCATGCACCCGTCACGCCGCTCAAGGCCAGCTGTGTGCACAGGTGCAG 44464
QY 352 GluAspThrIleLysValAlaValAspArgSerCysAspIleGlyAsnPheGluLeuLeu 371
Db 44463 GAGGACACCATCAAGGTGGCCGTGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTG 44404
QY 372 TTPArg----- 373
Db 44403 TGGAGGAGGTGAGCCACGCCGCCCTCTGGGACTTTTGTGGCCGCCGCCATGTGGATG 44344
QY 373 ----- 373
Db 44343 CTTTCCACCTGGTCACTGGGCGAGGTGGAGCTGCAGGCCGCCAGTCAGCGGGGTCTA 44284
QY 373 ----- 373
Db 44283 TGGAGGACACCATCTCTGTGTGACCATGCTCGTCCCGTGTAGGGGCCCGCAGGCCCTGC 44224
QY 373 ----- 373
Db 44223 CTGCTCTCCCATGGGCTCAGTGCCTCGTCCGCCACCGAAGGGGCTACTCCAGCGCTG 44164
QY 373 ----- 373
Db 44163 CGGGCGAGCGGCCCGCTCACTCAGGCTCAGTGCCTGTGTCAGCTTTCTCCTCTTTG 44104
QY 373 ----- 373
Db 44103 TTCCAGCCTCACCCAGTCTGTGACTGTGTGTAGAGTGGGGCTCCATGCCCGCAGCCTC 44044
QY 373 ----- 373
Db 44043 CAGGCCACCTATGTGCTCTTTGCTTGGCTGCGAGTGCCTCTCCACTCTGAGGATCT 43984
QY 373 ----- 373
Db 43983 TTCGGGGCTCCCTGTGCCGTGGGATGTGTGGAGTCATTTGGGAAGGTGGCCATGCC 43924
QY 373 ----- 373
Db 43923 TCTCTTTGGGCTCTCTGTGGCTAAACTCGGGAAGAGCAGCTGAGTCCCTGGGTGT 43864
QY 373 ----- 373
Db 43863 GGGGCCCCAGGCTGACCCCAAGCCTCAGGAAGCCTTGGTTCGCCCGAGGTTCTCCAG 43804
QY 373 ----- 373
Db 43803 AGCCTGGGGGCTCTCTGGGGGTCTGCTGGTGGGACCCAGGTGTTCACCTTCACTCTCTGAG 43744
```

QY 373 ----- 373

Db 43743 TCAGGTGGGGTCTGTGTCCCGCAGCAGCCCTCCTCTGTGCACAGACCCCTCTCTGCGAGGACCC 43684

QY 374 ----- 374

Db 43683 CCCCTATCCCGCCACACGCCAGCACCACCCAACTGTCTTCTCCACACAGCCGCTGTGA 43624

QY 378 uProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaArg 398

Db 43623 GCCGCCCCCATTCAGCGGGTCCGACCTCTGCGTGGCGGCGTCACTGTGAGGAGCCAG 43564

QY 398 gArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnPr 418

Db 43563 GAGGCGAGTGTCTGCCCTCTGCAACCTCAAGGCTCGGCTCGCTGTGAGACGCGCAGCC 43504

QY 418 oleuLysAlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAl 438

Db 43503 GCTGAAGGCACGGGGCCCTCGGCCATCCAGACCTGCCAGGACCCCAATCACCAGC 43444

QY 438 aLeuGlnArgAspLeuGlyLeuLysGluGlyLysGlyLeuProLeuAlaLeuAlaPr 458

Db 43443 TCTCCAGCGGACTTGGGACTTGAAGGAAGAGAGGGCTCCCTCTGGCTTGTGTCACC 43384

QY 458 oleuArgGlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaG 478

Db 43383 CTTAAGGGGGCAGCCGAGAGCGGTGGAGCCGCACAGCCACCCCAAGCTGCTGG 43324

QY 478 yLysValGluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGl 498

Db 43323 GAAGGTGGAGTCCCGGCTCTCCCTCTGTCGACAGTCAGGACAGTCAAGGCCCCAACACCGG 43264

QY 498 yValProValAlaGlnProAlaLysSerTTPAspProAsnGlnLeuAsnAlaHisProle 518

Db 43263 TGTCCCGTACCCAGCCGCCCAAAAGCTGGGATCCAAACAGCTAAATGCGACCCCT 43204

QY 518 uGluProValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGl 538

Db 43203 GGAGCTGTGTGTCGGGGCTTGAAGACAGACAGAGGGCGCTGCGCTCCGCCCGCGAGG 43144

QY 538 yLysGlySer 541

Db 43143 AAAAGGTAAC 43134

RESULT 12

US-10-615-659-3

; Sequence 3, Application US/10615659

; Publication No. US20040157234A1

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN

; FILE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42

; FILE REFERENCE: D0283 NP

; CURRENT APPLICATION NUMBER: US/10/615,659

; PRIOR FILING DATE: 2003-07-09

; PRIOR APPLICATION NUMBER: U.S. 60/394,725

; PRIOR FILING DATE: 2002-07-09

; NUMBER OF SEQ ID NOS: 102

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 3

; LENGTH: 726

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(726)

US-10-615-659-3

Alignment Scores:

Pred. No.: 3.31e-125 Length: 726

Score: 1299.00 Matches: 242

Percent Similarity: 100.0% Conservative: 0

Best Local Similarity: 100.0% Mismatches: 0
Query Match: 45.5% Indels: 0
DB: 7 Gaps: 0

US-10-635-977-2 (1-541) x US-10-615-659-3 (1-726)

QY 133 IleAspGlyLeuArgAsnIleTrpIleLeuLysProAlaAlaLysSerArgGlyArgAsp 152
DB 1 ATTGACGGGCTCCGGAACATCTGGATTATAAAGCCCGCGCAAGTCCCGGGCCGAGAC 60

QY 153 IleValCysMetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLeu 172
DB 61 ATAGTGTGATGACCGTGTGGAGGATCCTGGAGCTGGCAGCTGCAGACCACTCTCT 120

QY 173 SerArgAspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuLysCysAsp 192
DB 121 TCCAGGGACACACAGTGGGTGTCCAGAGTACATCGAGCGCGCTGCATCTGTGAC 180

QY 193 ThrLysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrp 212
DB 181 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGGACTGGAACTGCATCTGG 240

QY 213 PheTyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAsp 232
DB 241 TTCTACAAGGAGAGTACTTGGCGGTCTCAACTCAGCGCTTCTCCCTGGACAGCTGG 300

QY 233 SerAlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArg 252
DB 301 AGCGCCATCCACTGTGCAACACGCGTCCAGAGTACTTGAAGATGATGTGGCGCG 360

QY 253 SerProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluThrLeuGln 272
DB 361 AGCCCCCTGCTGCCCGCACACATGTGGACCGACAGCTGGAAGTACTTCCAGGAGTAC 420

QY 273 ArgGlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIle 292
DB 421 CGCAGGCGCGTGGCGCGTGTGGGCGAGCTTCACTACCCCGTCCATGAAGAAGGCCATC 480

QY 293 AlaHisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeu 312
DB 481 GCCCACCCATGAAGTGGCGCGAGGACCATCGTGAGCTTCCAGAGGAGTGCAGGAG 660

QY 313 TyrGlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluLeuAsnSer 332
DB 541 TACGGGCTGACTTCGTCTTGGAGGACTTTCAGGCGCTTGGCTGTGATCGAGATCAATTCC 600

QY 333 SerProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGlu 352
DB 601 AGCCCCACCATGACCCGCTCACGCCGCTCACGCCGAGCTGTGTGCACAGGTCAGGAG 660

QY 353 AspThrIleLysValAlaValAspArgSerCysAspIleGlyAsnPheGluLeuLeuTrp 372
DB 661 GACACCATCAAGGTGGCGGTGGACCGCGAGCTGTGACATCGGCACTTCGAGCTCCTGTGG 720

QY 373 ArgGln 374
DB 721 AGGCAG 726

RESULT 13
US-10-635-977-3
; Sequence 3, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09

; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(726)
US-10-635-977-3

Alignment Scores:
Pred. No.: 3,31e-125 Length: 726
Score: 1299.00 Matches: 242
Percent Similarity: 100.0% Conservatives: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 45.5% Indels: 0
DB: 7 Gaps: 0

US-10-635-977-2 (1-541) x US-10-635-977-3 (1-726)

QY 133 IleAspGlyLeuArgAsnIleTrpIleLeuLysProAlaAlaLysSerArgGlyArgAsp 152
DB 1 ATTGACGGGCTCCGGAACATCTGGATTATAAAGCCCGCGCAAGTCCCGGGCCGAGAC 60

QY 153 IleValCysMetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLeu 172
DB 61 ATAGTGTGATGACCGTGTGGAGGATCCTGGAGCTGGCAGCTGCAGACCACTCTCT 120

QY 173 SerArgAspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuLysCysAsp 192
DB 121 TCCAGGGACACACAGTGGGTGTCCAGAGTACATCGAGCGCGCTGCATCTGTGAC 180

QY 193 ThrLysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrp 212
DB 181 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGGACTGGAACTGCATCTGG 240

QY 213 PheTyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAsp 232
DB 241 TTCTACAAGGAGAGTACTTGGCGGTCTCAACTCAGCGCTTCTCCCTGGACAGCTGG 300

QY 233 SerAlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArg 252
DB 301 AGCGCCATCCACTGTGCAACACGCGTCCAGAGTACTTGAAGATGATGTGGCGCG 360

QY 253 SerProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluThrLeuGln 272
DB 361 AGCCCCCTGCTGCCCGCACACATGTGGACCGACAGCTGGAAGTACTTCCAGGAGTAC 420

QY 273 ArgGlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIle 292
DB 421 CGCAGGCGCGTGGCGCGTGTGGGCGAGCTTCACTACCCCGTCCATGAAGAAGGCCATC 480

QY 293 AlaHisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeu 312
DB 481 GCCCACCCATGAAGTGGCGCGAGGACCATCGTGAGCTTCCAGAGGAGTGCAGGAG 540

QY 313 TyrGlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluLeuAsnSer 332
DB 541 TACGGGCTGACTTCGTCTTGGAGGACTTTCAGGCGCTTGGCTGTGATCGAGATCAATTCC 600

QY 333 SerProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGlu 352
DB 601 AGCCCCACCATGACCCGCTCACGCCGCTCACGCCGAGCTGTGTGCACAGGTCAGGAG 660

QY 353 AspThrIleLysValAlaValAspArgSerCysAspIleGlyAsnPheGluLeuLeuTrp 372
DB 661 GACACCATCAAGGTGGCGGTGGACCGCGAGCTGTGACATCGGCACTTCGAGCTCCTGTGG 720

QY 373 ArgGln 374
DB 721 AGGCAG 726

RESULT 14

```
US-10-275-595A-39
; Sequence 39, Application US/10275595A
; Publication No. US20040078804A1
; GENERAL INFORMATION:
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: LU, Dying, Aina M.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: LAL, Preeti
; APPLICANT: YAO, Monique G.
; APPLICANT: BANDMAN, Olga
; APPLICANT: BURFORD, Neil
; APPLICANT: BATRA, Sajeev
; APPLICANT: KEARNEY, Liam
; APPLICANT: POLICKY, Jennifer L.
; TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0772 USN
; CURRENT APPLICATION NUMBER: US/10/275,595A
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/201,960
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/202,729
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 60/209,705
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 60/210,149
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: US 60/213,215
; PRIOR FILING DATE: 2000-06-21
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PERL Program
; SEQ ID NO 39
; LENGTH: 2380
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 2156553CB1
US-10-275-595A-39

Alignment Scores:
Pred. No.: 1.15e-100 Length: 2380
Score: 1070.50 Matches: 250
Percent Similarity: 54.8% Conservative: 81
Best Local Similarity: 41.4% Mismatches: 198
Query Match: 37.5% Indels: 76
DB: 7 Gaps: 12

US-10-635-977-2 (1-541) x US-10-275-595A-39 (1-2380)

QY 2 AlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerSerArgSer 21
DB 285 GCCCGCAACGTTCTCAAGCTGGTGTGAAGTCTGAG-----TGAAGTCA 329

QY 22 LysProArgAspGlnArgGluAlaGlySerSerAspLeuSerSerArgGlnAspAla 41
DB 330 TACCTATTTCAGGCGAGTAGAGAGAGCGCTTCAGAGACAGCAGCCCAAGAAACAGGAG 389

QY 42 GluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLys 61
DB 390 AAAAACCCAGTGTG-----GTGTCCCGAGAGTTGTGGATGAAGCTCTGTGT 437

QY 62 ValCysGlnAlaTyrrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAspAla 81
DB 438 GCGTGGAGGAGTACCTTAGCAACTTGGCCACATGGACATCGACAAGGACCTGGAGGCC 497

QY 82 ValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrrSerLeuVal 101
DB 498 CCGTGTACCTCACCCCGAGGGCTGGTCCCTCTTCCTCCAGCGCTACTACCAAGTGGTC 557
```

```
QY 102 HisGlyAspAlaPheIleSerAsnSerArgAsnTyrrPheSerGlnCysGlnAlaLeuLeu 121
DB 558 CACGAAGGGCGAGAACTCAGGCACCTCGACACTCAGGTCCAGCGCTGTGAGGACATCCTG 617

QY 122 AsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIle 141
DB 618 CAGCAGCTGCGAGCGCGTGTACCCAGATAGACATGGAGGGGATCGCAACATCTGGATC 677

QY 142 IleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGlu 161
DB 678 GTGAAGCCAGGAGCCAAAGTCCCGTGGAGGAGCATGTGTCATGGACCATCTGGAGGAG 737

QY 162 IleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValGln 181
DB 738 ATGCTGAAGCTGGTGAACGGCAACCCCGTGGTGTGATGAAGACCGCAAGTGGGTGGCAG 797

QY 182 LysTyrrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPhe 201
DB 798 AAGTATATTGAGCGCGCCCTCTCTATCTTTGGCACCAAGTTTGACCTCAGACAGTGGTTC 857

QY 202 LeuValThrAspTrpAsnProLeuThrIleTrpPheTyrrLysGluSerTyrrLeuArgPhe 221
DB 858 CTGTTAACTGACTGGAAACCCACTTACCGTGTGGTTCTACCGGACAGCTATATCCGCTTT 917

QY 222 SerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsnAla 241
DB 918 TCCAGCGACGCTTCTCTCCCTGAAGAACCTCGGACAACTCAGTGCACCTGTGCAACAACTCC 977

QY 242 ValGlnLysTyrrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMet 261
DB 978 ATCCAGAAAGCACCTGGGAACCTCATGCCATCGGCATCCACTGCTTCCGCCACAGCAACATG 1037

QY 262 TrpThrSerThrArgPheGlnGluTrpLeuGlnArgGlnGlyArgGlyAlaValTrpGly 281
DB 1038 TGGTCTAGCAGAGTTCCAGGCCCACTCGAGGAGATGGTGGCCCAAAATGCTTGGTCC 1097

QY 282 SerValIleTyrrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAsp 301
DB 1098 ACCATCATCGTCTGGCATGAAGATGCTGTGATCCACGCACTTTCAGACCTCCCGAGGAC 1157

QY 302 HisValGluProArgLysAsnSerPheGluLeuTyrrGlyAlaAspPheValLeuGlyArg 321
DB 1158 ACCGTGCAGTGTCCGAAGGCCAGCTTTGAGCTCTATGGCGCTGACTTCTGTGTCGGGAG 1217

QY 322 AspPheArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThrPro 341
DB 1218 GACTTCCAGCCCTCGCTGATTGAGATCAACCCAGGCCCAAGATGGCACTTCCACAGCA 1277

QY 342 ValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaVal----- 359
DB 1278 GTCACTGCCCGGCTCTGTGTGGCTGCAAGCTGACACCTCGCGCTGTGTCATTGACCGG 1337

QY 360 -----AspArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValVal 377
DB 1338 ATGCTGGACCGCAACTGTGACACAGGAGCCCTTTCAGCTCATCTATAAGCAGCTCTGTG 1397

QY 378 GluProProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgArg--- 396
DB 1398 GAGTGGCTCAATATGTGGGCATCCGGCTCTCTGTAGAGGGCTTCACCATCAAGAAGGCC 1457

QY 397 -----AlaArgArgGlnValLeuProValCysAsnLeu----- 407
DB 1458 ATGGCGATGTCTATCGCGGATGGGGGTTCGCCCGACAGTCCCTCTCTGTGACCCAGCGA 1517

QY 408 -----LysAlaSer-AlaSerLeuLeuAspAlaGlnProLeu---LysAl 421
DB 1518 GGCTCTGGGGAGGCAAGGACTCGGGATCCCTCCACAGCTCAGCTTCTTAGGAAGGCG 1577

QY 421 aArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln--- 440
DB 1578 ACTGGGCGCAGGAGCCCTGGGCGACAGTGAGAAAGCGAGTCTCCACTGCCACCACCTTCAGCC 1637
```

```
QY 441 -ArgAspLeuGlyLeuLysGluGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuAr 460
Db 1638 CCGGAAAGGGGAGAAAGGCAAGCGAAGGGCCACAGCCCTGGTGTGCCCCCAATCTC 1697
QY 460 gGlyAlaAla-----GluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAl 477
Db 1698 TGGGAGTGGGATGCCCCAGCAGCAGGATGGGCTGATTTTC-ACCATGACCTTTTCTAG 1756
QY 477 aGlyLysValGlu-----LeuProAlaCysProCysArgHisValAspSerGlnAl 494
Db 1757 TGGGGACAGCAACCCACCACTTGAACAGATTGCCACTGAGTCCGAGCAAGAACCCCCAGGC 1816
QY 494 aProAnThrGlyValProValAlaGln-----ProAlaLys 506
Db 1817 CTGGGTAAAGACATTCCCCCAAAACACCCGAGTGTTCAAAGCGGATTTATTCTCTCT 1876
QY 506 sSerTApProAsnGlnLeuAsnAlaHisPro----- 517
Db 1877 CCAAGCCCTCCCAACCACTGGATCAGCCACCCCAAGAGCCACCAAGTAGCAAGTA 1936
QY 517 ----- 517
Db 1937 AAAGCCACTACTCACAAGTATTGTTTAAATAATACACAGCCAAATTTAGCTGGGCGCGTG 1996
QY 518 -LeuGluProValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGln 537
Db 1997 GTGTGAGCCTGTGTCCAGGCTACTCGCGAGGCTAATGAGGAT---CGTTGAGCCCGG 2053
QY 537 yGlyLysGly 540
Db 2054 AGGTCAAGGC 2063
```

RESULT 15

```
US-10-102-524-1765
; Sequence 1765, Application US/10102524
; Publication No. US20030109434A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Gaiger, Alexander
; APPLICANT: Gordon, Brian
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; THERAPY AND DIAGNOSIS OF KIDNEY CANCER
; FILE REFERENCE: 210121.572
; CURRENT APPLICATION NUMBER: US/10/102,524
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 1863
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1765
; LENGTH: 2538
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-102-524-1765
```

Alignment Scores:

```
Pred. No.: 2,32e-95 Length: 2538
Score: 1020.00 Matches: 220
Percent Similarity: 55.5% Conservative: 69
Best Local Similarity: 42.2% Mismatches: 151
Query Match: 35.7% Indels: 81
DB: 5 Gaps: 12
```

US-10-635-977-2 (1-541) x US-10-102-524-1765 (1-2538)

```
QY 54 GlnLeuValAspLeuAlaCysLysValCysGlnAlaThrLeuGlyGlnLeuGluHisGlu 73
Db 835 GAGTTTGTGATGAAGCTCTGTGTGCTGCGGAGGAGTACCTTAGCAACTTTGGCCCATG 894
QY 74 AspLeuAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluThrGluAspLeu 93
Db 895 GACATCGACAGGACCTGGAGGCCCGCTGTACCTACCCCGGAGGCTGTGCTCTCTTC 954
```

```
QY 94 ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
Db 955 CTCCAGCGCTACTACCAAGTGTTCACGAGGGGAGAACTCAGGACACCTCCACACTCAG 1014
QY 114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspLeu 133
Db 1015 GTCCAGCGCTGTGAGGACATCTCGCAGCAGCTGCAGGCGGTGGTACCCAGATAGACATG 1074
QY 134 AspGlyLeuArgAsnIleThrIleLeuLeuProAlaAlaLysSerArgGlyArgAspLeu 153
Db 1075 GAAGGGGATCGCAACATCTGGATCTGAAGCAGGAGGCCAAGTCCGCGGAGGAGCATC 1134
QY 154 ValCysMetAspArgValGluGluLeuAlaAlaAspHisProLeuSer 173
Db 1135 ATGTGATGAGCACCTGGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGGTGATG 1194
QY 174 ArgAspAsnLysTyrValGlnLysTyrIleGluThrProLeuLeuIleCysAspThr 193
Db 1195 AAGGACGGCAAGTGGGTGGTGCAGAAATATATTGAGCGGCCCTCTCTCATCTTTGCGACC 1254
QY 194 LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
Db 1255 AAGTTTGTACCTCAGACAGTGGTTCTCTGTTAACTGACTGGAAACCACTTACCGTGTGTT 1314
QY 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
Db 1315 TACCGGAGCAGCTATATCCGCTTTTCCACGAGCCCTTCTCTCTGAGAACCTTGGACAAC 1374
QY 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
Db 1375 TCAGTCACCTGTGCAACAACCTCCATCCAGAGAACCTGGAGAACTCATGCCATCGGCAT 1434
QY 254 ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnLysLeuGlnArg 273
Db 1435 CCACTGCTTCCGCCACACAAACATGTGTCTAGCCAGAGTTCCAGGCCCACTGCGAGGAG 1494
QY 274 GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla 293
Db 1495 ATGGGTGCCCAAAATGCTTGGTCCACCATCATCTGCTGGCATGAAGATGCTGTGATC 1554
QY 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
Db 1555 CAGGCATTCAGACCTCCCGAGCAGACCGTGCAGTGTGCGAAGGCCAGCTTGTAGCTCTAT 1614
QY 314 GlyAlaAspPheValLeuGlyArgAspPheArgProTrpIleLeuGluLeuAsnSerSer 333
Db 1615 GCGCTGACTTGTGTTCGGGAGGACTTCCAGCCCTGTGATTCAGATCAACACCGCAGC 1674
QY 334 ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353
Db 1675 CCAAGATGGCACCTCCACAGCAGTCACTGCGCGCTCTGTGCTGGCGTGCAGCTGAC 1734
QY 354 ThrIleLysValAlaVal-----AspArgSerCysAspIleGlyAsnPheGlu 369
Db 1735 ACCCTGCGGTGTCTATTGACCGGATGCTGGACCGCAACTGTGACAGAGGCTTTGAG 1794
QY 370 LeuLeuTrpArgGlnProVal-----ValGluPro-ProProPheSerGlySerAspLeuCy 388
Db 1795 CTCATCTATAAGCAGCCCGCTCACCACTTCCCGAGCTCCACACCAAGGC----- 1843
QY 388 sValAlaGlyValSerValArgAlaArgGlnValLeuProValCysAsnLeuLys 408
Db 1843 ----- 1843
QY 408 sAlaSerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArgGlyProSerAlaMetPr 428
Db 1844 -----CCAGCTGCTCTTCCCCCATGTACTCCGACACCCAGGGCCAGGCTCTCAGACGACAGC 1899
QY 428 oAspProAlaGlnGlyProProSerProAlaLeuGlnArgAspLeuGlyLysGluGln 448
Db 1900 ACAGCAAGCTGTGGGCACTTAAGGCCCTGTGCGACCAACAGCA-----GGCCTTGGAGGATC 1956
QY 448 uLys-----GlyLeuProLeuAlaLeuAlaPro---LeuArgGlyAla----- 462
```

```
Db      1957 TACCCACGGCTAAGGCTTCATTCTCCCTCCACGAACCTTGATTTCAAGGTGGCACCCA 2016
Qy      |||||
Qy      463 -----AlaGluSerGlyAlaAlaGlnProThrArgThr----- 474
Db      |||::: |||||
Db      2017 GCATCCTGAAGCCAAGAAAGGTGGCTCGACCTGTGACTCACACCCAGTGGACAGTGCT 2076
Qy      |||::: |||||
Qy      475 -----LysAlaAlaGlyLysValGluLeuProAlaCysProCysAr 488
Db      |||::: |||||
Db      2077 GAGCACGGGTCAAGGCTGGAGGGCACAGGCAGAGGGCAGCTCCAGGCTGG----- 2128
Qy      488 gHisValAspSerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTr 508
Db      |||::: |||||
Db      2129 -----CTGGCACCCCAAGGGAAGAGCTG 2151
Qy      508 pAspProAsnGlnLeuAsn-----AlaHisProLeuGluProValLe 522
Db      |||::: |||||
Db      2152 GTCTCCCTCAGAAGCCCTTCCTCCACAGACTTCTGATCATCTCCCTCTCTCCCTCCT 2211
Qy      522 uArgGlyLeuLysThrAlaGluGlyAlaLeuArgPro-----ProProGlyGl 538
Db      |||::: |||||
Db      2212 T-----TCACACCGAGGCTCCTGCTCTCCTGTGCTTCGAGGCCCCCAGCTGG 2259
Qy      538 Y 538
Db      2260 A 2260
```

Search completed: April 4, 2006, 11:54:48
Job time : 1232 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2006, 09:38:21 ; Search time 288 Seconds
(without alignments)
3339.100 Million cell updates/sec

Title: US-10-635-977-2

Perfect score: 2854

Sequence: 1 MASSILKVVVSHQSCSRSR.....LRGLKTAEGALRPPPGKGKS 541

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/abs/ABSSWEB spool/US10635977/runat_04042006.103535.10086/app_query.fasta_1
-DB=Issued Patents NA -QPMF=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=Bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-HOST=abs02h -USER=US10635977 @CGN 1 1.193 @runat 04042006.103535.10086
-NCPU=6 -ICPU=3 -NO MMAP -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WREN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/1 COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5 COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/H COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
7: /cgn2_6/ptodata/1/ina/PP COMB.seq.*
8: /cgn2_6/ptodata/1/ina/RE COMB.seq.*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1020	35.7	3001	3	US-09-620-312D-494
2	305	10.7	5314	3	US-09-620-312D-138
3	302	10.6	2624	3	US-10-104-047-1688
4	271.5	9.5	2074	3	US-10-104-047-1401
5	228.5	8.0	1980	3	US-09-270-767-15138
6	210	7.4	2111	3	US-10-104-047-1651
7	205.5	7.2	2716	3	US-09-799-451-295
8	166	5.8	3394	3	US-09-814-915A-56
9	154.5	5.4	783	3	US-09-270-767-31557

10	154.5	5.4	1525	3	US-09-270-767-15222	Sequence 15222, A
11	153	5.4	30135	3	US-09-902-540-1249	Sequence 1249, Ap
12	151	5.3	780	3	US-09-248-796A-500	Sequence 500, App
13	132	4.6	3027	3	US-09-252-991A-12848	Sequence 12848, A
14	129.5	4.5	1212	3	US-09-252-991A-12504	Sequence 12504, A
C 15	129.5	4.5	1251	3	US-09-252-991A-12970	Sequence 12970, A
C 16	129.5	4.5	7419	3	US-09-252-991A-481	Sequence 481, App
C 17	129.5	4.5	7449	3	US-09-252-991A-396	Sequence 396, App
C 18	129	4.5	1188	2	US-08-926-253-1	Sequence 1, Appli
C 19	129	4.5	1188	2	US-08-926-253-3	Sequence 1, Appli
C 20	129	4.5	1188	3	US-09-120-074-1	Sequence 1, Appli
C 21	129	4.5	1188	3	US-09-120-074-3	Sequence 3, Appli
C 22	129	4.5	1297	3	US-09-270-767-15012	Sequence 15012, A
C 23	129	4.5	1297	3	US-09-270-767-15012	Sequence 1273, Ap
C 24	127	4.4	2534	3	US-09-906-776B-1	Sequence 1, Appli
C 25	127	4.4	2534	3	US-09-923-922-1	Sequence 1, Appli
C 26	127	4.4	3765	3	US-09-949-016-5816	Sequence 5816, Ap
C 27	127	4.4	3801	3	US-09-949-016-907	Sequence 907, App
C 28	127	4.4	23951	3	US-09-902-540-1245	Sequence 1245, Ap
C 29	127	4.4	27903	3	US-09-902-540-1235	Sequence 1235, Ap
C 30	126.5	4.4	12194	3	US-09-902-540-1091	Sequence 1091, Ap
C 31	126	4.4	2715	3	US-09-252-991A-12548	Sequence 12548, A
C 32	126	4.4	5173	3	US-09-949-016-11822	Sequence 11822, A
C 33	126	4.4	5173	3	US-09-949-016-13985	Sequence 13985, A
C 34	125.5	4.4	3342	3	US-09-252-991A-12644	Sequence 12644, A
C 35	125.5	4.4	5467	3	US-09-902-540-703	Sequence 703, App
C 36	125.5	4.4	6244	2	US-08-076-726-15	Sequence 15, Appli
C 37	125.5	4.4	6244	2	US-08-260-452-8	Sequence 8, Appli
C 38	125.5	4.4	6244	2	US-08-481-970-8	Sequence 8, Appli
C 39	125.5	4.4	6244	2	US-08-897-719-8	Sequence 8, Appli
C 40	125.5	4.4	6244	3	US-09-163-269-8	Sequence 8, Appli
C 41	125.5	4.4	6244	3	US-09-281-674-8	Sequence 8, Appli
C 42	125.5	4.4	6244	3	US-09-777-3178-8	Sequence 8, Appli
C 43	125.5	4.4	7305	3	US-09-902-540-961	Sequence 961, App
C 44	125.5	4.4	61158	3	US-09-949-016-15041	Sequence 15041, A
C 45	125.5	4.4	75431	3	US-09-949-016-15122	Sequence 15122, A

ALIGNMENTS

RESULT 1

US-09-620-312D-494
; Sequence 494, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenchua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620.312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0

```
; SEQ ID NO 494
; LENGTH: 3001
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1361)..(2419)
US-09-620-312D-494

Alignment Scores:
Pred. No.: 7,34e-84 Length: 3001
Score: 1020.00 Matches: 220
Percent Similarity: 55.5% Conservative: 69
Best Local Similarity: 42.2% Mismatches: 151
Query Match: 35.7% Indels: 81
DB: 3 Gaps: 12

US-10-635-977-2 (1-541) x US-09-620-312D-494 (1-3001)

QY 54 GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGlu 73
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1304 GAGTTTGTGATGAAGCTCTGTGTGCTGCGAGGAGTACCTTAGCAACTTTGGCCCATG 1363
QY 74 AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTyrGluAspLeu 93
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1364 GACATCGACAAAGACCTGGAGGCCCGCTGTACTACCCCGGAGGCTGCTCCCTCTTC 1423
QY 94 ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1424 CTCAGCGCTACTACCAAGTGGTCCACGAAGGGGAGAACTCAGGCACCTCGACACTCAG 1483
QY 114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1484 GTCCAGCGCTGTGAGGACATCTCGACGAGCTGCAGCGCTGTGTGTACCCAGATAGACATG 1543
QY 134 AspGlyLeuArgAsnIleTyrPheLeuProAlaAlaLysSerArgGlyArgAspIle 153
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1544 GAAGGGGATCGCAACATCTGGATCTGTAAGCCAGGAGCAAGTCCCGGACGAGGATC 1603
QY 154 ValCysMetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLeuSer 173
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1604 ATGTGCATGACCACTGGAGGAGATGCTGAAGTGTGTGAACGCAACCCCGTGTGATG 1663
QY 174 ArgAspAsnLysTyrValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThr 193
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1664 AAGCAGCGCAAGTGGTGGTGCAGAGTATATTGAGCGGCCCTCTCTCATCTTTGGCACC 1723
QY 194 LysPheAspIleArgGlnTyrPheLeuValThrAspTyrAsnProLeuThrIleTyrPhe 213
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1724 AAGTTTGACCTCAGACAGTGGTCTCTGGTAACCTGACTGGAACCCACTTACCGTGTGTTT 1783
QY 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1784 TACCGCGACAGCTATATCCGCTTTTCCACGAGGCCCTCTCCCTGGAAGAACCTGGACAAC 1843
QY 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1844 TCAGTGACCTGTGCAACAACTCCATCCAGAAGCACCTGGAGAACTCATGCGCATCGCAT 1903
QY 254 ProLeuLeuProAlaHisAsnMetTyrThrSerThrArgPheGlnGluTyrLeuGlnArg 273
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1904 CCAGTGTTCGCGCAGACAAATGTGTGTGTAGCAGAGGTTCAGAGCCACCTCGAGAG 1963
QY 274 GlnGlyArgGlyAlaValTyrGlySerValIleTyrProSerMetLysLysAlaIleAla 293
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1964 ATGGGTGCCCCAAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2023
QY 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
2024 CACGCACTTCAGACCTCCAGGACACCGTGCAGTGTGCGAAGCCAGCTTTGACCTCTAT 2083
QY 314 GlyAlaAspPheValLeuGlyArgAspPheArgProTyrLeuIleGluIleAsnSerSer 333
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
```

RESULT 2

```
US-09-620-312D-138
; Sequence 138, Application US/09620312D
; Patent No. 659662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing-A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
```

; APPLICANT: Wang, Zhiwei
 ; APPLICANT: John Tillinghast
 ; APPLICANT: Dmanac, Radoje T.
 ; TITLE OF INVENTION: No. 656962el Nucleic Acids and
 ; FILE REFERENCE: Polypeptides
 ; CURRENT APPLICATION NUMBER: US/09/620,312D
 ; CURRENT FILING DATE: 2000-07-19
 ; PRIOR APPLICATION NUMBER: 09/552,317
 ; PRIOR FILING DATE: 2000-04-25
 ; PRIOR APPLICATION NUMBER: 09/488,725
 ; PRIOR FILING DATE: 2000-01-21
 ; NUMBER OF SEQ ID NOS: 1105
 ; SOFTWARE: pt_FL_genes Version 1.0
 ; SEQ ID NO 138
 ; LENGTH: 5314
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (531)..(4376)
 ; US-09-620-312D-138

Alignment Scores:
 Pred. No.: 1,89e-17 Length: 5314
 Score: 305.00 Matches: 89
 Percent Similarity: 49.3% Conservative: 50
 Best Local Similarity: 31.6% Mismatches: 93
 Query Match: 10.7% Indels: 51
 DB: 3 Gaps: 11

US-10-635-977-2 (1-541) x US-09-620-312D-138 (1-5314)

QY 137 ArgAsnileTrpIlelelyysProAlaAlaLysSerArgGlyArgAspileValCysMet 156
 Db 1050 CGGGACCTTGATAGTAAACAGTCAGTCATCTTCAAGGGCGGGCGGTCTACCTGATC 1109
 QY 157 AspArgValGluGluLeuLeuAlaAlaAspHisProLeuSerArgAspAsn 176
 Db 1110 AACATCCAAACAGATC-----TCCCTGGAAGAG 1139
 QY 177 LysTrpValValGlnlystFyleGluThrProLeuLeuLeuCysAspThrlyspheAsp 196
 Db 1140 AACATTTGGTCTCCGGTTACATTAACACCCCTGCTCATAGATGATTTCAAGTTTGAC 1199
 QY 197 IleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGlu 216
 Db 1200 GTGGCCCTCTAGTCTGCTGACTTCCCTATGATCTCTTGTCATCTATCTATGAA 1259
 QY 217 SerTyrLeuArgPheSerThrGlnArgPhe-----SerLeuAspLysLeuAspSer 233
 Db 1260 GGATTGGCTAGGTTGGCACTGTGGATATGATCAAGGAGCCCAAGACATTCCGGAACCA 1319
 QY 234 AlaIleHisLeuCysAsnAsnAlaValGlnlyys-----TyrLeuLysAsnAsp 249
 Db 1320 TTCATGTCATTCGACAACTACAGTGTCAACAGAAAGAGTGGAGATTACGTGTTGTGAC 1379
 QY 250 ValGlyArgSerProLeuProAlaHis---AsnMetTrpThrSerThrArgPheGln 268
 Db 1380 -----GATCCAGAGTGGAGGATTATGGAACAATGGAGCATGCTATGCTT 1430
 QY 269 GluTyrLeuGlnArgGlnGlyArgGlyAla----- 278
 Db 1431 AGGTACCTGAAACAAGAGGAGAGATACACCCGATTTGATGGCCCATGTAGAACACCTG 1490
 QY 279 ValTrpGlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysVal 298
 Db 1491 ATCATTAAGACTATAATC-----TCTGCTGAACCTAGCTATGCTACTGCTGTAACACC 1544
 QY 299 AlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheVal 318
 Db 1545 TTTGTTCTCAT-----CGCAGCAGTGTGTTTGAATCTATGGCTTTGACGTGCTC 1595

QY 319 LeuGlyArgAspPheArgProTrpLeuLeuGluLeuAenSerSerProThrMetHisPro 338
 Db 1596 ATAGATTCTACTCTGAAGCCATGGTGTGGAGTGAATCTCTCTCTCTTTGGCCTGT 1655
 QY 339 SerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIle----- 355
 Db 1656 GATGCGCTCTGGACCTAAAGATTAAAGCCATGATGATTTCAGATAT-GTTCACTCTTGT 1714
 QY 356 LysValAlaValAspArgSerCysAspileGly-----Asn-Ph 368
 Db 1715 AGGATTGTGTGCCAAGATCTCTGCCAGCGGGCATCACTCGGCCAATTTATCCACCTT 1774
 QY 368 eGluLeuLeuTrpArgGlnProValValGluProPro-----ProPheSerGlySe 385
 Db 1775 CGAGTCTTCCAGCGCAACCTTTCAGAAACCTCAGCGTTCGCGTCCACTCTCTGCCAG 1834
 QY 385 rAsp 386
 Db 1835 TGAT 1838
 RESULT 3
 US-10-104-047-1688
 ; Sequence 1688, Application US/10104047
 ; Patent No. 6943241
 ; GENERAL INFORMATION:
 ; APPLICANT: HELIX RESEARCH INSTITUTE
 ; TITLE OF INVENTION: No. 6943241el full length cDNA
 ; FILE REFERENCE: H1-A0105
 ; CURRENT APPLICATION NUMBER: US/10/104,047
 ; CURRENT FILING DATE: 2002-03-25
 ; PRIOR APPLICATION NUMBER:
 ; PRIOR FILING DATE:
 ; NUMBER OF SEQ ID NOS: 4096
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1688
 ; LENGTH: 2624
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-104-047-1688
 Alignment Scores:
 Pred. No.: 1.19e-17 Length: 2624
 Score: 302.00 Matches: 132
 Percent Similarity: 37.7% Conservative: 76
 Best Local Similarity: 24.0% Mismatches: 200
 Query Match: 10.6% Indels: 144
 DB: 3 Gaps: 23
 US-10-635-977-2 (1-541) x US-10-104-047-1688 (1-2624)
 QY 82 ValGluAspLeuThrGluAlaGluTrpGlu----- 91
 Db 447 GTGAAAGTGAAGGAGGAGGAGTGGATTTCTACTGTTGACGTGAGTGGTCCGG 506
 QY 92 -----AspLeuThr-----GlnGlnTyr 97
 Db 507 GAGAACTTCGACCACCTACATCGATGAACATGTGCGGATCAGTCTTCGGAACCA 566
 QY 98 TyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCys 117
 Db 567 TATGAGCTGACCCCGAAGAACTATCATGTGTGAAGAACCTGAAGCGGTTCGGAAGCAGCTG 626
 QY 118 GlnAlaLeuLeuAsnArgIleThrSerValAsn-----ProGlnThr----- 131
 Db 627 GAGCGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 686
 QY 132 -----AspIleAspGlyLeuArgAsn-----IleTrp 140
 Db 687 ATGCTTTCGAGTACCACCTGTTTGTAGAGGAGTTTCGAAAAAACCAGGAATCACCTGG 746
 QY 141 IleIleIysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
 Db 747 ATCATGAAGCCTGTAGCCCGGTCTCAAGGGAAGGCATCTTCTCTTCCGTTAGGCTGAAG 806

```

471 ProThrArgThrLysAlaAlaGlyLysValGluLeuProAlaCysProCysArgHisVal 490
    |||:::  ::::|  |||  :::  |||  |||  |||  |||  |||  |||  |||  |||
1787 CCITCTGCCTTCAGAGCATGAGGACTTCTCCCACTCATCCCTGCTACCTGAGGCACACAA 1846
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
491 AspSerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTrpAspPro 510
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1847 CCCAGACAG-----GTCCAGGCGTGGACTCCA 1873
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
511 AsnGlnLeuAsnAlaHisProLeuGluProValLeuArgGlyLeuLysThrAlaGluGly 530
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1874 GGTTCCTCA-----TTTCCAGAAGAAATGCGATTAGCTTTTTTT 1912
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
531 AlaLeuArgProPro---ProGlyGlyLysGly 540
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1913 GGCCAGCACCCACCCTGCCCTGCCAGGAGGAGAGGGG 1945
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

RESULT 4
US-10-104-047-1401
; Sequence 1401, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1401
; LENGTH: 2074
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1401

Alignment Scores:
Pred. No.: 5,42e-15 Length: 2074
Score: 271.50 Matches: 107
Percent Similarity: 36.6% Conservative: 59
Best Local Similarity: 23.6% Mismatches: 149
Query Match: 9.5% Indels: 139
DB: 3 Gaps: 15

US-10-635-977-2 (1-541) x US-10-104-047-1401 (1-2074)

QY 2 AlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerSerArgSer 21
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 14 GCCTTCGGCGCATTTTCAGCTGCGTGCAGC-----TCTGCACAGACACCCACAGAGGC 67
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 22 LysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerSerArgGln----- 39
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 68 CACCCTCGAACCAACGCGCCCAATGAGAGGGCGGACCTGTGTCTCCACACAAAGCCAG 127
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 40 -----AspAla 41
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 128 GCGCTGGGATCTTTGAGAACCAACCAACCCCGAGCCTTTACCTTTAACATTCCTCGAGGCA 187
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 42 GluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValaAspIleAlaCysLys 61
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 188 AACCACTAGACACCGCGCTGCAGGCGCTGGAGCAAGGCTACAGGACGAGGTGTTTCC 247
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 62 Val----- 62
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 248 ATCCCTCCAGCGGAGCGCGCCCAACCAACCACTGGAGAGAAGAAAAACCTCATTTG 307
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 63 -----CysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAsp 76
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 308 ATGGCGGAAGATGAACCTTCAGGGGCGCTCTTGAAGCGCGCTGGTTTTCGCGTGTGACGAG 367
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 77 ThrSerAlaAspAlaValGluAsp---LeuThrGluAlaGluTrpGluAspLeuThrGln 95
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 368 ACCACCCGGCGTGTGGTGCAAGATGTCCTCGGAGAGGGGGTGAATAGTTTGATAAG 427
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

QY	161	GluIleuLeuGluLeuAlaAlaAlaAspHisProLeuSerArgAspAsnLys-----	177
Db	807	GAACATCGTGACATGG---AGGAAGGACACCAAGAAGCTCTGACGACCAAGAAATGATATATT	863
QY	178	-----TrpValValGlnLysbTyrlleGluThrProLeuLeuIleCysAspThr	193
Db	864	CCCGTGGAGAACTATGTGGCTCAGCGTTACATTTGAAATCTTACTGATAGGAGCGCG	923
QY	194	LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe	213
Db	924	AAAGTTTGACCTGGCTCTATGTGCTGGTGATGTG-----GTG	962
QY	214	TyrLysGluSerTyrIleuArgPheSerThrGlnArgPheSerIleuAspLysLeuAspSer	233
Db	963	TTTGCTGAATGCTCGTGTGTGTGGCGCACGGAGA-----CAG	1001
QY	234	AlaIleHisIleuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer	253
Db	1002	GATGTTCACTCACCACTGGCTGTGTGCAAAA-----	1034
QY	254	ProLeuLeuProAlaHisAsn-----MetTrpThrSerThrArgPheGln	268
Db	1035	---ACATCTCCGACTACCAACCAAGAGGGCTGCAAGTGGACGCTGCAGCGCTTCGG	1091
QY	269	GluTyrIleu---GlnArgGlnGlyArgGlyAlaValTrpGlySerValIleTyrProSer	287
Db	1092	CAGTACCTGGCGTCCAAACACGGGGCCCGAGGCAGTG-----GAGACACTCTTCAGGGAC	1145
QY	288	MetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAspHisValGluProArgLys	307
Db	1146	ATCGACAACATCTTTGTCAAAAGCCCTGCAGAGTGTGCAGAAGTGATCATCAGTGACAAG	1205
QY	308	AsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeu	327
Db	1206	CACGTCTTCAGCTGTACGCTATGACATCTCATCGACCAGGACCTCAAGCGCTGCCTC	1265
QY	328	IleGluIleAsnSerSerProThrMetHisProSerThrProValThrAlaGlnIleuCys	347
Db	1266	CTGGAGTCAATGCTGCCCTCACTGCACAGCCAGGACGCCAGGAAGACTATGAGCTCAAG	1325
QY	348	AlaGlnValGlnGluAspThrIleLysValAla-----ValAspArg	361
Db	1326	ACCTGCCTCTGGAAAGACACCTGCATGTTGTGGACATGAAGCGAGGCTCACGGGAAG	1385
QY	362	SerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValValGlu-----	378
Db	1386	GAGAAGCGAGTCCGGGGCTTTGACCTCATGTGGAATGATGCCCTGTTAGCAGAGGAG	1445
QY	379	-----ProProProPheSerGlySerAspIleuCysValAlaGlyValSerValArgArg	396
Db	1446	GGGGCTCTCGA-CCTGTGGGAATGGGAAACTTTGTGCACCAACACATCTCGGTGCGT	1504
QY	397	AlaArgArg-----	399
Db	1505	CAACGATCGGAAGAAACAACTGAGCGAGCTCTTCTGCTCCCTTCAAGTTCAGAAGAAAGC	1564
QY	400	---GlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnPro	418
Db	1565	TTCCAGTTGATCCCGACGCTGCCAGGAGGAAATCAGCCTTAGC-----AGGCCT	1612
QY	419	Leu-----LysAlaArgGlyProSerAlaMetProAspPro	430
Db	1613	TGTGTTTGAGATGGAGCTGAGCCAGAAACCGGAAACCTCTGGGCTCATGGCGGACGA	1672
QY	431	AlaGlnGlyProProSerProAlaLeuGlnArgAspLeuGlyLeuLysGluGlySerGly	450
Db	1673	GCTTCGGAGTTGAGTGTCAAGCCCGAAGGGAAGACTTT---TGGAGGAAAAATACTGGC	1729
QY	451	LeuProLeuAlaLeuLeuAlaProIleuArgGlyAlaAlaGluSerGlyAlaAlaGln	470
Db	1730	AAATCCAAGAGTGAACCGCGACTT---TAGGAAGCCAGAAAGCCAGGCTACTCTAGA	1786

```
QY 96 GlnTyr-----TyrSerLeuValHisGlyAspAlaPheIleSerAsnSer 110
|||
Db 428 CAGGACGACAAACCGGAGGACTGGAACTGTACTG-GAGGACATCTCTTCGGAATGAC 486
|||
QY 111 ArgAsnTyrPheSerGlnCys-----
|||
Db 487 CGRACA-----CAACAGTGTAAACCGTGGCAGCAGCTAAACCACCCCTGGAAACAC 540
|||
QY 118 GlnAlaLeuAsnArgIleThrSerValAsnPro-----
|||
Db 541 CAAGCTTACCAGGAAGACTGTTTGGCCAAACACCTGAAGCACATGAGGAGGATGTATGG 600
|||
QY 129 -----
Db 601 CACTTCCCTGTACCACTTCATCCCTGACGTTGCTGTCATGCCCAATGACTATACCAAGTT 660
|||
QY 130 -----GlnThrAsp-IleAspGlyLeuArgAsnIleTrpIleI 142
|||
Db 661 CGTGGCTGAATCTTTCAGGAGGAGGAGATGCTGGCCACCAAGCATAGCTATGATTGG 720
|||
QY 142 eLysProAlaAlaLysSerArgGlyAspIleValCysMetAspArgValGluGluI 162
|||
Db 721 CAAGCCTGCTGAGTTATCTCGTGGGAGGGGATACTAATTTTCAGTGACTTTAAAGACTT 780
|||
QY 162 eLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValGlnIly 182
|||
Db 781 CATC-----TTTGATGATATGTACATAGTCAGAA 810
|||
QY 182 sTyrlleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPhe 202
|||
Db 811 ATATATCTCCATCTCTTACTTATTTGGCAGATATAAATGTATGATCTCCGATCTATGTTTG 870
|||
QY 202 uValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPhe 222
|||
Db 871 TGTACTGCTTAAAGCTTTGACCATTTATGTTTATCAGAAAGGTTGGTTCGTTTGC 930
|||
QY 222 rThrGlnArgPheSerLeuAspLysLeu---AspSerAlaIleHisLeuCysAsnAl 241
|||
Db 931 CACGGAAGAGTTGACCTCAGTAATTTGCCAAACAATATATGCCCATTTGACCAACAGCAG 990
|||
QY 241 aValGlnLys-----TyrLeuLysAsnAspValGlyArgSerProle 255
|||
Db 991 CATCAATAAATCCGGGGCTCTTATGAGAAGATCAAGAAGTGTATGGTCATGGT- 1045
|||
QY 255 uLeuProAlaHisAsnMetTrpThrArgPheGlnGluTyrLeuGlnArgGlnG 275
|||
Db 1046 -----TGTAATGACGCTCAGCAGATTTTCTTCTACCTTCGTAGC- 1087
|||
QY 275 yArgGlyAlaValTrp-----GlySerValIleTyrProSerMetLysLysAla 292
|||
Db 1088 -----TGGGATGTGGACGATCTGCTTTTGTGGAAGAAAATCCACCGCATGT 1134
|||
QY 292 eAlaHisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGlu 312
|||
Db 1135 TATCTTCACCATCTCGCATTCGCCATCTCTCCCTTTGCTGCCAATTCGTTTGGACT 1194
|||
QY 312 uTyrGlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluIleAsnSe 332
|||
Db 1195 CTTTGGGTTTGATATTTTGTATTTGATGATGACNACTTGAACCATGCTTTTAGAGGTCACTA 1254
|||
QY 332 rSerPro-----ThrMetHisProSerThrProVal 342
|||
Db 1255 CAGCCCGACCTTGACCTTGATGTTGTTCAACAGATGTG 1291
|||
```

RESULT 5

```
US-09-270-767-15138
; Sequence 15138, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
```

```
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15138
; LENGTH: 1980
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15138
```

Alignment Scores:

```
Pred. No.: 4,718-11 Length: 1980
Score: 228.50 Matches: 83
Percent Similarity: 40.7% Conservatives: 53
Best Local Similarity: 24.9% Mismatches: 110
Query Match: 8.0% Indels: 88
DB: 3 Gaps: 11
```

US-10-635-977-2 (1-541) x US-09-270-767-15138 (1-1980)

```
QY 179 ValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArg 198
|||
Db 1 GTGGTGTGAAATATATAGTGGATCGCTGTCGTCATTGATGGTCACAAATGTGATTGAGG 60
|||
QY 199 GlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyr 218
|||
Db 61 GTATACGTTCTAGTCACCTTCCTTCGATCCCTTAATTTATTTATTTATATGAAGGGCAAT 120
|||
QY 219 LeuArgPheSerThrGlnArgPheSer-----LeuAspLysLeu---AspSerAlaIle 235
|||
Db 121 GTGAGATTGGCCACAGTCAAGTACGACGAGGATCCGATAACCTGTGGAAACCCGTGCATG 180
|||
QY 236 HisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsn----- 248
|||
Db 181 CATCTCTGCAATTACAGCATTAACAATACCATCTGATTACATAGGAGCTCCGATGCT 240
|||
QY 249 -----AspValGlyArgSerProLeuLeuProAlaHisAsnMetTrpThrSerThr 265
|||
Db 241 CAGGATGAGGATGTAGGC-----CATAAGTGGACGTTGTGTC 276
|||
QY 266 ArgPheGlnGluTyrLeuGlnArgGlnGly-----ArgGlyAlaValTrpGly--- 281
|||
Db 277 GCTCTTTTAAAGCACTGAAACTCCAGAGTGTGTATACACGTCAGCTAATGCTGAATATT 336
|||
QY 282 ---SerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln 300
|||
Db 337 GAGGATTTAATTTAAGCTGTGTACCTGTGCTCAATCTATTATTTCGGCCTGCAGG 396
|||
QY 301 AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly 320
|||
Db 397 ATGTTCTGTTCCCAATGGAATAAATGCTTTGAGCTGTATGGATTTCGATATATAATAGAC 456
|||
QY 321 ArgAspPheArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThr 340
|||
Db 457 AACGCACATAAGCCATGCTACTTGGAAATCAATCTATCCGCTCAATGGGTGTGGATAGT 516
|||
QY 341 ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValaAsp 360
|||
Db 517 CTTTGGACACTAAGTT----- 534
|||
QY 361 ArgSerCysAspIleGlyAsn-PheGluLeuLeuTrpArg---GlnProValValGluPro 379
|||
Db 535 AAGTCATGCTGATGGCAGACCTATTGACGTTGTGGGCATTTCAGGCTATAGTC----- 589
|||
QY 379 oProPheSerGlySerAspLeuCysValAlaGlyValSerValArgArgAlaArgAr 399
|||
Db 589 ----- 589
|||
QY 399 gGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProle 419
|||
Db 590 -----CCGAATGAATCTCATTATACCCAGAAATGGTCGCGTTCCGCGAGCTCA 639
|||
QY 419 uLysAlaArgGlyProSerAlaMetProAspProAlaGlnGlyProSerProAlaLe 439
|||
```

Db 640 AGCTGCAAGGATAGCAACTTTTCCAGTACCTCCAAAAGCAAGAAGT----- 691
Qy 439 uGlnArgAspLeuGlyLeuGlyGluGlyLeuGlyGlyLeuProLeuAlaLeuAlaProLe 459
Db 692 -----CGAAGAAAAAAGGAGCTGCCATTAACTTAATCTTAACCGGC 732
Qy 459 uArgGlyAlaAla-----GluSerGlyG1 467
Db 733 GAAGAGCAGCGCATATTGCGCAATGCAACGCTACAGTATTCTCGTGTGAGGATTCGTG 792
Qy 467 yAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal 480
Db 793 CGCATCTTTCCACAGATGACTCAATGCACGCGTATGGTA 832

RESULT 6

US-10-104-047-1651
; Sequence 1651, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIORITY FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1651
; LENGTH: 2111
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1651

Alignment Scores:
Pred. No.: 2,66e-09 Length: 2111
Score: 210.00 Matches: 146
Percent Similarity: 32.7% Conservative: 53
Best Local Similarity: 24.0% Mismatches: 202
Query Match: 7.4% Indels: 210
DB: 3 Gaps: 31

US-10-635-977-2 (1-541) x US-10-104-047-1651 (1-2111)

Qy 6 LeuLysTrpValValSerHisGlnSerCysSerArgSerArgSerLysProArgAsp 25
Db 532 CTGAGTGG-----TGTGAGTCAAGAGCCGAGACAGCTACGGCAGC 573
Qy 26 GlnArgGluAlaGlySerSerAspLeuSerSerArgGlnAspAla----- 41
Db 574 TTCGGGAAGG-AGAGCAGCTGCTGTACCAGCTTCCCAACAACAAAGCTCTCCACCACAA 632
Qy 42 GluAsnAlaGlu-AlalysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLy 61
Db 633 GATCGGCTGCTCAGCACCTTCTCGGGA---CGGGCACGGGCCCATGAGCGGGCCAGCAA 689
Qy 61 sValCysGlnAlaTyrlLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAspAl 81
Db 690 GGTGCCGGGGGGTCCAGGCCAGCTGGAAAG-----GACCGACAGCGGCCCGC 740
Qy 81 aValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrlSerLeuVa 101
Db 741 CCTGGAGGACCTC-----CCGTGG----- 759
Qy 101 lHisGlyAspAlaPheIleSerAsnSerArgAsnTyrlPheSerGlnCysGlnAlaLeuLe 121
Db 760 -----ACAAGCCAGGATACCTCAGGCCACAGAGGGTCTCG-- 795
Qy 121 uAsnArgIleThrSerValAsnProGlnThr-----AspIleAspGlyLeuArg-- 137
Db 796 ----AGATGGAAGAGTTTTTCCAGAGACCTTACCGCTGAGCTCAACACACGAGAGAGA 851

Qy 138 -----AsnIleTrpIleIleLysProAlaAlaLysSe 148
Db 852 GGCCTTTTTCACCTTGTGTGATGAACCCAGATATGGATCTGCAAGCCACAGCCTCCAA 911
Qy 148 rArgGlyArgAspIleValCysMetAspArgValGluGluIleLeuGluLeuAlaAla-- 167
Db 912 CCAGGGCAAGGATCTTCTCTCGTCCGGAACCCAGGAGGAAGTTCGCCCTCGAGGCCAA 971
Qy 168 -----AlaAspHisProLeuSerArgAspAsnLysTrp----- 178
Db 972 GACCCGAGCATGGAGGACGACCCCATCCACCAAGACGCCGTTCGGGGGGCTCAGGC 1031
Qy 179 ----ValValGlnLysTyrlleGluThrProLeuLeuIleCysAspThrLysPheAspIl 197
Db 1032 GCGGTGTGTGACAGGTATCATCAGAACCCGCTGTGTGGACGGAGAGAAGTTTGACGT 1091
Qy 197 eArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrlLysGluSe 217
Db 1092 GCGCTCTACCTGCTCATTTGCTGTGACACACACCTCATATGATC---TTCTTTGGCCACCG 1148
Qy 217 rTyrlLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLe 237
Db 1149 CTATGCTCGCTCACCCCTTAGCCTTTACGACCCCATTCACGCGACCTCGGGCGGCACCT 1208
Qy 237 uCysAsnAsnAlaValGlnLysTyrlLeuLysAsnAspValGlyArgSerPro----- 254
Db 1209 GACCAACCAAGTTTCATGCAGAAG-----AAGAGCCTCTGTACAT 1247
Qy 255 -LeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrlLeuGlnArgGl 274
Db 1248 GCTGCTGAAGGAGACACACGCTGTGGAGC-----ATGGAACATCTCAACCGCTA 1295
Qy 274 nGlyArgGlyAlaValTrpGlySerValIleTyrlProSerMetLysLysAlaIleAlaHi 294
Db 1296 CATCAGTGACACGTTCTGG-----AA 1316
Qy 294 sAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyrlG1 314
Db 1317 GGCCTGGGGCTCGCCCAAGGAC----- 1338
Qy 314 yAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluIleAsnSerSerPr 334
Db 1339 -----TGGGTCTTC----- 1347
Qy 334 oThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspTh 354
Db 1348 -----ACCACCTCAAGGTGGCTCCACTGTGC----- 1374
Qy 354 rIleLysValAlaValAspArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGl 374
Db 1375 -----CC 1376
Qy 374 nProValValGluProProPheSerGlySerAspLeuCysValAlaGlyValSerVa 394
Db 1377 TCCAGTCTGGAGTGAGATCCC---TCGGGGCGGGG---TGTGTGTGAGGTGGGCAC 1430
Qy 394 lArg-----ArgAlaArgArgGlnValLeuProValCy 405
Db 1431 CAGGCACACAGATGTCCGTGGCGTGGCGCGCTCGCTGAAGTGTGACCTGACCGCTG 1490
Qy 405 sAsnLeuLys-----AlaSerAlaSerLeuLeuAspAlaGlnProLeuLy 420
Db 1491 TGAACCAAAACCTTCCAGCGTCTGTCTCATTTAGCTGGCAGTGTCTCCAGCAGC 1550
Qy 420 sAlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGl 440
Db 1551 CAGCAAGGCCCGACGACGAAAGCCAGTCGGGGTCTGTCTGGCAGCAGTCCCGGGCAG 1610
Qy 440 nArgAspLeuGlyLeuLysGluGlyLysGlyLeuProLeuAla-----LeuLe 456
Db 1611 CCTCGCGTAGGTCAGGGGACAGCTCGGCTTCAGTGACAGCCACCATGTGAAGAGTCTT 1670
Qy 456 uAlaProLeuArgGlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAl 476

Db 1671 TGTTCCTTCAGATCTCCACCTGGCAGGAGG-----CCACAGCAGCCCCCGG 1721
Qy 476 aLaGlyLysValGlu-LeuProAlaCysProCysArg----- 488
Db 1722 GCAGGTGCTGGAGCAGCAGCTGCCGTGCGAGGCCGAGCTCTCCCTAGCTAAG 1781
Qy 489 --HisValAspSerGlnAlaProAsn-----ThrGlyValProValA 502
Db 1782 GCATGTGCG-GTGGAGTGTCTAAATTTTGCAGGGTTCCTTCTGTGGGGGTACCTGTGG 1840
Qy 502 la-----GlnProAlaLysSer-----TTPAspP 510
Db 1841 GGTACTTCAACAGCCCTAGCAGCAAGGCCCTTGAGCAGCGCGGTGCAACTGGGATA 1900
Qy 510 roAsnGlnLeuAsnAlaHisProLeuGluProValLeuArgGlyLeuLysThrAlaGluG 530
Db 1901 AAAAC-----GGGGCTTGGCGGAGGACTTTTATCTG-----TCTT 1936
Qy 530 lyAlaLeuArgProProProGly 537
Db 1937 GGTACCCCGCCGAGGCCCGGA 1959

RESULT 7

US-09-799-451-295
; Sequence 295, Application US/09799451
; Patent No. 6783969
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Ma, Yunqing
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Dunrui
; APPLICANT: Yang, Yonghong
; APPLICANT: Wehrman, Tom
; APPLICANT: Ghosh, Reena
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6783969el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/09/799,451
; CURRENT FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 948
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 295
; LENGTH: 2716
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (277) .. (2442)
US-09-799-451-295

Alignment Scores:
Pred. No.: 1.02e-08 Length: 2716
Score: 205.50 Matches: 78
Percent Similarity: 42.2% Conservative: 49
Best Local Similarity: 25.9% Mismatches: 116
Query Match: 7.2% Indels: 59
DB: 3 Gaps: 11

US-10-635-977-2 (1-541) x US-09-799-451-295 (1-2716)

Qy 106 PheIleSerAsnSerArgAsnTyrPheSerGlnCys----- 117

Db 80 TATGTAGTTTGAATAGAAATTAATTTTCTTATTATGTCCTCAATTTACAGAAGGATCAAGT 139
Qy 118 -----GlnAlaLeuLeuAsnArgIle 124
Db 140 CTCGGCCTCTGGATTATACCTTTTGTTCCTCGAAGCTTGGATCTTCTCTGCTGAATATACTC 199
Qy 125 ThrSerValAsnProGlnThrAspIleAspGlyLeuArg-----AsnIleTrpIleIle 142
Db 200 AATTCCAA-AATTATGTGAAGAAATTGAAGAAAAACCGAAGCAGAAAAACTTTTATAGTG 258
Qy 143 LysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGluIle 162
Db 259 AAACAGAGTAATGGTCAATGGGTCTATGGGATTTCTTGTATAGAAATAGTGCAC----- 312
Qy 163 LeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValGlnLys 182
Db 313 ---AACTTCCATCTCAGATCAT-----TTGATTGTTCAAGAA 348
Qy 183 TyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPheLeu 202
Db 349 TACATTGAAAAGCCTTCTTAATGAAGGTTACAAGTTTGACTTACGAATTTATATTCTG 408
Qy 203 ValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPheSer 222
Db 409 GTTACATCGTGTGATCCACTAAAAATATTCTCTACCATGATGGGCTTGTGCGAATGGGT 468
Qy 223 ThrGlnArgPhe-----SerLeuAspLysLeuAspSerAlaIleHis 236
Db 469 ACAGAGAGTACATCCACCTAATGAGTCCAAATTTGACCAGTTA-----TACATGAT 522
Qy 237 LeuCysAsnAsnAlaValGlnLys-----TyrLeuLysAsnAspValGlyArgSer 253
Db 523 CTGACAAACTACTCCGTGAACAAGCATAAATGAGCATTTTGAACGGGATGAAACTGAGAAC 582
Qy 254 ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArg 273
Db 583 AAAGCAGCAAAACGTTCCATCAAAATGG-----TTTACAGAATTCCTTCAAGCA 630
Qy 274 GlnGlyArgGlyAlaVal-----TrpGlySerValIleTyrProSerMetLysLysAla 291
Db 631 AATCAACATGATGTTGCTAAAGTTTTGGAGTGAATATTTCAGAAATGCTGTTAAGACCCTG 690
Qy 292 IleAla-----HisAlaMetLysValAlaGlnAspHisValGluPro 305
Db 691 ATTGTAGCAGAACCTCATGTCTCTCATGCTATCGAATGTGTAGACCTGGTCAACTCCA 750
Qy 306 ArgLysAsnSer-----PheGluLeuTyrGlyAlaAspPheValLeuGlyArgAspPhe 323
Db 751 GGAAGCGAAAGTGTCTGCTTTGAAAGTCTCGGATTTGATATTTTGTGGATAGAAAACTA 810
Qy 324 ArgProTrpLeuIleGluLeuAsnSerSerProThrMetHisProSerThrProValThr 343
Db 811 AAGCCATGGCTTCTGGAGATTAAACCGAGCCCAAGCTTTTGGAACTGATCAGAAAAATAGAC 870
Qy 344 AlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAspArgSerCys 363
Db 871 TATGATGTAAAAAGGGGAGTCTGCTAAATCGTTGAAGTACTATAAACATAAGGACCAGT 930
Qy 364 Asp 364
Db 931 GAC 933

RESULT 8

US-09-814-915A-56
; Sequence 56, Application US/09814915A
; Patent No. 6750015
; GENERAL INFORMATION:
; APPLICANT: Horwitz, Kathryn
; APPLICANT: Richter, Jennifer
; TITLE OF INVENTION: Progesterone Receptor-Regulated Gene Expression and Methods Rel
; TITLE OF INVENTION: Theredo
; FILE REFERENCE: 2848-39

; CURRENT APPLICATION NUMBER: US/09/814,915A
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: 60/214,870
; PRIOR FILING DATE: 2000-06-28
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56
; LENGTH: 3394
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-915A-56

Alignment Scores:
Pred. No.: 6,43e-05 Length: 3394
Score: 166.00 Matches: 132
Percent Similarity: 32.6% Conservative: 64
Best Local Similarity: 22.0% Mismatches: 206
Query Match: 5.8% Indels: 202
DB: 3 Gaps: 25

US-10-635-977-2 (1-541) x US-09-814-915A-56 (1-3394)

```
QY      8  TrpValValSerHisGlnSerCysSerArgSerSerArgSer----- 21
DB      874 TGTGTGACCCCGGCGGCACATCTTCAGGCTTACAGGAGCTGCAGCGAGTGCAGCGTGGCGGCGCA 933

QY      22 -----LysProArgAspGlnArgGluGluAlaGlySerSer 33
DB      934 GCCTCACCCACCGCGCTTCACCTCACCCAGAGTGAGCGGCGCGCATCTCTTCA 993

QY      34 AspleuSerSerArgGlnAspAlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGly 53
DB      994 ACTTCTCACACTTCAAGGACTACAGGAACTCA-----GCCAGGAG-AGGCAGGC 1043

QY      54 GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGlu 73
DB      1044 GTGCTG-----CTGAACAGTTCCCTCGCGAG 1070

QY      74 AspleuAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu 93
DB      1071 AACCTGTGACTGTCAAGGAGTGCCTGCGCTCCATCGCGCGCGGCGAGTGGCGCC-- 1127

QY      94 ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
DB      1128 -----GAGGGGCCACCTGGTGGTCCCGCAACCTTCAACCTG 1163

QY      114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
DB      1164 CGCACTGAG-----CTGCCCCAGTTTGTCACTACTTCCAGCAGCGGGAAGG 1211

QY      134 AspGlyLeuArgAsnIleTrpIleLeuProAlaAlaLysSerArgGlyArgAspIle 153
DB      1212 TGGGGCGAGGACCAACCACTGGATCTGCAAGCCCTGGAACCTGGCGCGCAGCTGGAC-- 1268

QY      154 ValCysMetAspArgValGluGluLeuLeuGluAlaAlaAspHisProLeuSer 173
DB      1269 -----ACCCAGCTCACCAAGAG-CCTGCACAGCATCATCCGGCACCG 1309

QY      174 ArgAspAsnLysTrp-ValValGlnLysTyrIleGluThrProLeuLeu----- 189
DB      1310 AGAGCAGCCCCCAAGTTGTGTCCAGTATCATCGAAGTCCCGTGTGTCTTCCGAGA 1369

QY      190 ----IleCysAspThrLysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnPr 208
DB      1370 AGACGTGGGAAGGTCAAGTTCGACATCCGCTACATCGTGTCTGGGTCAAGTAGGCC 1429

QY      208 oLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLe 228
DB      1430 CCTACGTTGTTCGTGTATGATGTGTTGTGCTGCGGTTCTCAACCGCGGCTTTGCACT 1489

QY      228 uAspLysLeuAsp----- 232
DB      1490 CAACGACCTGGATGACTACGAGAGGACACTTTCACGGTCAATGAACTATGACCCCGGATGGT 1549
```

RESULT 9

US-09-270-767-31557

; Sequence 31557, Application US/09270767

; Patent No. 6703491

```
QY      233 ----SerAlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGI 251
DB      1550 GCTGAAGCAGGTGCAC---TGTGAAGAGTTTATCCCGAGTTTGAGAAG----- 1595

QY      251 yArgSerProLeuLeuProAlaHisAsnMetTrpThrSerThrArg----- 266
DB      1596 -CAATACCCAGAAATTCCC-----TGGACGGACGTCACGGCTGAGATCTTCCG 1642

QY      267 ----PheGlnGluTyrLeuGln---ArgGlnGlyArgGlyAlaValTrpGlySerValII 284
DB      1643 GGCCTTCACGGAGCTGTTCCAGTGGCTGTGTCAGCAGCCACCCCTGGGCTCTGCGCA 1702

QY      284 eTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAspHisValGI 304
DB      1703 CTACCCCTCA----- 1712

QY      304 uProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArgAsp----- 322
DB      1713 -----TCCCGGGCCATGTATGCCGTCGACCTCATGCTGAAGTGGGACAACGG 1759

QY      323 -----PheArgProTrpLeuIleGluIleAsn-SerSerProThrM 336
DB      1760 CCCAGATGGAAGCGGCTGATGCAGCCGAGATCTCTGGAGGTGAACCTTCAACCCGACTG 1819

QY      336 etHis-----ProSerThrProValThrAlaGlnLeuCysAlaG 349
DB      1820 TGAGCAGGCTGAGGATGACACCCACCTTCTTCAAGGAGTCTTCAAGACCTTGT----- 1875

QY      349 lnValGlnGluAspThrIleLysValAlaValAspArgSerCysAspIleGlyAsnPheG 369
DB      1875 ----- 1875

QY      369 luLeuLeuTrpArgGlnProValValGluProProPheSerGlySerAspLeuCysV 389
DB      1876 ----TTCGGACACCGCGTGGCTGCCACGTTACCT----- 1908

QY      389 alAlaGlyValSerValArgArgAlaArgGlnValLeuProValCysAsnLeuLysA 409
DB      1909 -----GCCTGTCTTAGGCACTCGCTGTCCCAAACTGTGTGGGGC-AGGA 1956

QY      409 laSerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArgGlyProSerAlaMetProA 429
DB      1957 TTCAACCTCAGTTCTTGTAGCTGCTTCTGCAAGGCC-----CCCATGTCTCTCCCA 2010

QY      429 spProAlaGlnGly-----ProProSerProAlaLeuGlnA 441
DB      2011 CACCGGCTCTGGGCGATAGCCTCAGCCCGCAGGCTCTGTCTCGGAGCCATCTCCCGGC 2070

QY      441 rgAspLeuGlyLeuLysGluGluLysGlyLeuProLeuAlaLeuAlaProLeuArg 461
DB      2071 GCCAC-----ACTCGGAGGACAGCATCTCTCTCTCACT 2106

QY      461 lyAlaAlaGluSerGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysValG 481
DB      2107 GTGGTCAGAGCAGGACAGTGTGGTCTCCC-----AGGCGTG 2145

QY      481 luLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro 501
DB      2146 AGCACCAACCCCGCCCTGCGCTCAC-----CCCTCACCCACCATCTGTGCACCTGATGAG 2199

QY      501 alAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGluPro 521
DB      2200 TCTCAGTTAGCAAGGCTTGTCTCTGGCATGGGAATTTGTTTC---CTGGCTGCTG 2256

QY      521 alLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGlyLysGly 540
DB      2257 TGTT-----CCAGGGGCTGCTGGG 2276
```


7729 GACTTCGGGACCCCGAGCGCCCTTCGCAAGAGATTCTCCCTCGCCCTGGGCC 7788
Db
48 LeuArgGly---LeuProGlyGlnLeu-ValAspIleAlaCysLys---ValCysGlnAl 65
Qy
7789 GCTCAGGATGAGCGCTGCGCGCTTCGATTGACAAGGCTTTCGAGAGGTCTTCCAAG 7848
Db
65 a-----TyrLeuGlyGlnLeuGl 71
Qy
7849 CGAGCGCTGCGGAGCGGGCCGAGCGCTCCAAAGTGGATGACTTCGACGGAGCTCCG 7908
Db
71 uHisGluAspIleAspThrSerAlaAsp-----AlaValGluAspLeuThrGl 87
Qy
7909 CAACAGAACCAGAAATACCCGCGGAGGAGTCCGCTTCGCCATCCAGACAAGCTGGC 7968
Db
87 uAlaGluTrpGluAspLeuThrGlnGlnTyrTrpSerLeuValHisGlyAspAlaPheIl 107
Qy
7969 ATTGGAGCGGAGCGACTG-----GAGCGCCATTGCGCGGCAACACTGACAGGTTTCAT 8022
Db
107 eSerAsnSerArgAsnTyrPheSerGlnCys----- 117
Qy
8023 CTCGAGGCCATCGCGTGGGTCAACCACTGCTACGACAAAGCGGCCCTCCACGGA 8082
Db
118 -----GlnAlaLeuAsnArgIleThr-SerValAsnProGlnTh 131
Qy
8083 GCGGACATCAGCAGCTGGCGCGCTTCGCGCGGACCTGATGAGCAAGATCCAAGAC 8142
Db
131 rAspIleAspGlyLeuArgAsnIleTrpIleLysProAlaAlaLysSerArgGlyAr 151
Qy
8143 GACCCGAGGAGCTCAAGTACGACCATCATCGAGCGCTCGCGGCAAGGTGCGGGAAC 8202
Db
151 gAspIleValCysMetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisPr 171
Qy
8203 GGACGCGGTGACGCGCGGAGAACATCGAAGTTTCATCCAGGACGCGTC----- 8251
Db
171 oLeuSerArgAspAsnLysTrpValValGlnLysTyrIleGlu-----Th 186
Qy
8252 -----AAGTGGTCAACCCAGTGTCTACGAGCAGACGAGCGCGGAGGCCAC 8295
Db
186 rProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrp----- 200
Qy
8296 GCCGCG-----GAGTGAAGGAGTGGCGCGCTTCGCCCCAGGA 8334
Db
201 -----PheLeuValThrAsp-- 205
Qy
8335 GCAGTTGAAGAAGACCCGGAATCTCCCGGAGCACCTCAAGTACGCCATCACCGACG 8394
Db
206 -----TrpAsnProLeuThrIleTrpPheTyrLy 215
Qy
8395 CATTCGCGCGAAGATGACGGGACCGGTGAGCGACCGCGTGNACATCGACAGTTTCAT 8454
Db
215 sGluSerTyrLeuArgPheSerThrGlnArgPhe---SerLeuAspLysLeuAspSerAl 234
Qy
8455 CGAGGACGGCTCAAGTGGTTCACCCAGTGTACGAGGGCGCAGGAGCGCAAGCCACCGC 8514
Db
234 aIleHisLeuCysAsn-----AsnAlaValGlnLysTyrLeuLysAsnAspValGlyAr 252
Qy
8515 CCGGGAACCTGGCGCACTGGCGCGACTTCGCCCGGAAGCACCAAGACCGGACCGGACAT 8574
Db
252 gSerPro-----LeuLeuProAlaHisAsnMe 261
Qy
8575 CACGCCGAGGCGCTCAAGTTGCGCATCCAGGACGCCATCCGCGTGAAGGCGAGCGGCAT 8634
Db
261 t-----TrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyAlaVa 279
Qy
8635 GGACGGCACGGGACGACCAACATCGACCACTTCATCCAG-----GACGCGGTCAA 8685
Db
279 lTrpGlySerValIleTyrProSerMetLys----- 289
Qy
8686 GTGGGTGTCTTCTGCTACGAGGGGCGAGGAGCGCGACGCCACCGCGGAGATGCGAGCG 8745
Db
290 ----LysAlaIleAlaHisAlaMetLysValAlaGlnAspHisValGluProArgLysAs 308
Qy
8746 CTGGCGCGCTTCGCCAACAGAGAGAACGAGACATCTCCCGGAGGAGCT 8805
Db
308 nSerPheGluLeuTyrGlyAla-----AspPheValLeuGlyArgAspPheArgPro-- 325
Qy
8806 CAAGTTTCGCCATCCAGGACCGCATGTCGCGACGAGAAGATGGGACCGCTCCGCCAG 8865
Db
326 -----TrpLeuIleGluLeuAsn----- 331
Qy
8866 CAATCGCATCGAAGACCATCCGGCGGCGCTACGGCTGGCTCATCTACCTCAACCGGCG 8925
Db
332 -----SerSerProThrMetHisProSer----- 339
Qy
8926 CCGGCCCTCGCAGCGCTCGAGCGCGACGCGCCACAGAGCTGCACGAATGGAGAAGTT 8985
Db
340 -----ThrPro---ValThrAlaGlnLeuCysAlaGlnValGl 351
Qy
8986 CGCCGAAAAGAGATTGAAGGAGCGCTGAGATGACGAGCGAGAGCTCCGGTACTACAT 9045
Db
351 nGluAspThrIleLysValAlaValAspArgSerCysAspIleGlyAsnPheGluLeuLe 371
Qy
9046 CTGGACAGCTCCGACCGCATGCGCAACAGTT--GACGGGCGCGCCCTCGCTCGCTCGC 9103
Db
371 uTrpArgGlnProValValGluProProPheSerGlySerAspLeuCysValAlaGl 391
Qy
9104 GTTCAGAGCGAGCGCAGGAAGGCCCAAT-GCCTCAGCTCGCGCTTGCTCAAGGCC-- 9160
Db
391 yValSerValArgArgAlaArg---ArgGlnValLeuProValCysAsnLeuLysAlase 410
Qy
9161 -----TTGAAGCGTTCGCGGCGACGAGCGCTTCGCCCCCGTGCACAGGATGGCCTC 9213
Db
410 rAlaSerLeuLeuAspAlaGlnPro-LeuLysAlaArgGlyProSerAlaMetProAspp 430
Qy
9214 C---TGAAGTTCGCGCGCGCGCTCGTGGAGCATCCGACCTGACGGTTACCGCTC 9270
Db
430 roAla-----G 432
Qy
9271 CAGCAACCCAGTCCCAAGGGCGGCTCTCCACTCGGTGCGGTCTCTCCGTC 9330
Db
432 lngly-----ProProSerProAlaLeuGlnArgAspLeuGlyLeuLysGluGlyLysG 450
Qy
9331 AGGGCGCGCTCGCGCAGCTCCGCGCCCATGTGTGAAGAGGGTCTGTCCCTTCATCAGGTGGGCTC 9390
Db
450 ly----- 450
Qy
9391 GATGCGTGTGGACAGCTCCGGACATCTCGACCGTCCCGGTGATGAATCCCGCTG 9450
Db
451 -----LeuProLeuAlaLeuAlaP 458
Qy
9451 GTGGCACCGCGCAGCGCTCTTCTGGAAGAGGGTCTGTCCCTTCATCAGGTGGGCTC 9510
Db
458 roLeuArgGlyAlaAlaGluSerGlyAlaAlaGlnProThrArgThrLysAlaAlaG 478
Qy
9511 CAACACGTCCGCGCGCGGCGGCGACGCGATGACCGGAGTACAGCGTCTGCTCGCTCCAG 9570
Db
478 lyLys-----ValGluLeuProAlaC 485
Qy
9571 GAGTGGCGGCAACACCGGGGGCCCTTGGCGGCCACTCGCGTCTATGATGAGCTTCTCTGCT 9630
Db
485 ysProCysArgHisValAspSerGlnAlaProAsnThrGlyVal----- 499
Qy
9631 CGCGGTGC-----AGTTGCTCTCCGATAGCGGTGACGCTCAGTCCAGGTC 9678
Db
500 -----ProValAlaGlnProAlaLysSerTrpAspProAsnGlnL 513
Qy
9679 CTCGGAGAAGCGCTGGCGACCTGTGCTCCAGGTGGGTGATGATGGCTTCCACCCGA-- 9737
Db
513 euAsnAlaHisProLeuGluProValLeuArgGlyLeu-----LysThrAlaGluGlyA 531
Qy
9738 -----ACGGGCCAGCTCCACCTTGGCGGAGTGGACGCTGACACCGCGGTTCGGACGGCC 9792
Db
531 laLeuArg-----ProProGlyGlyLysGly 540
Qy
9793 GGAGACCGCTTCCCATCAGCGTTCATCCGGGTTCGCGAGGCGGTGA 9839
Db

RESULT 12

US-09-248-796A-500

; Sequence 500, Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 500

; LENGTH: 780

; TYPE: DNA

; ORGANISM: Candida albicans

US-09-248-796A-500

Alignment Scores:

Pred. No.:	0.000159	Length:	780
Score:	151.00	Matches:	53
Percent Similarity:	44.0%	Conservative:	50
Best Local Similarity:	22.6%	Mismatches:	81
Query Match:	5.3%	Indels:	50
DB:	3	Gaps:	9

US-10-635-977-2 (1-541) x US-09-248-796A-500 (1-780)

Qy	132	AspIleAspGlyLeuArgAsnIleTrpIleIleLysProAlaAlaLysSerArgGlyArg	151
Db	58	GAATTAATAAAGAGAAAATGTCGATTTTGAACCTAGTAGTAGTAAAGGTCAA	117
Qy	152	AspIle-----ValCysMetAspArgValGluGluLeu-----	163
Db	118	GGTATAGAAATTTTAAACACGATCGATCAATTCGAAGAAATTTTCAACTCATTTGAAGAA	177
Qy	164	-----GluLeuAlaAlaAspHisPro-----Leu	172
Db	178	AACGATGAGAAATGAAGAAATGAGGAGGATGACGAGCAAGACGATAACGGTATAATTTTG	237
Qy	173	SerArgAspAsnLysTrpValValGlnLysTrpIleGluThrProLeuLeuIleCysAsp	192
Db	238	TCTCAACTTCGTCATTTATTGTTTCAGATATAGAGTAATCCCTTGTTACTTA-----	291
Qy	193	ThrLysPheAspIleArgGlnTrpPheLeu-----ValThrAspTrpAsnProLeu	209
Db	292	TCAAGATGATGACCATAGAATAATTCATTTACGTACTTATGTTGTTGTTGTTGTTG	351
Qy	210	ThrIleTrpPheTyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeu---	228
Db	352	AAAGTATTTGTTTATAGAAATGTTTTCACCTTATTTGCTGGTGAACCATATAAATTCGCT	411
Qy	229	-----AspLysLeuAspSerAlaIleHisLeuCysAsnAsnAlaValGlnLys	244
Db	412	GGTGACGAGATGAAGTAGTATCTTTAGCAGGACCTTTAACCACTGTTGTTTACAGGAA	471
Qy	245	TyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMetTrpThrSer	264
Db	472	-----AATGAGGATCCGTTGGTG-----	489
Qy	265	ThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTrpGlySerValIle	284
Db	490	-----GTACCATTTTGGAAATTAACAAGGTCTTTGCTGATAATGATAAAACATGTG	540
Qy	285	TyrProSerMet-----LysLysAlaIleAlaHisAlaMetLysValAlaGln	300
Db	541	TTTGACAAATTTGTCATATACAAAGAAATGTTTAAAGCTGCTCAAGAGTGTGACAAA	600
Qy	301	AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly	320

```

Db      601   ATGAATTTTCAACCAATAAATAAGCCATTGAGATATTGGTGTAAGATTCCTGGTGAAC 660
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      321   ArgaspPheArGProTrpLeulleCuluilehsnSerSerPro 334
          ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      661   AGTGATTTTCTCGGTAATCTATTAGAAGTTAAATTCATACCCT 702
          ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-09-252-991A-12848
; Sequence 12848, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12848
; LENGTH: 3027
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12848

```

Alignment Scores:					
Pred. No.:	0.0742	Length:	3027		
Score:	132.00	Matches:	149		
Percent Similarity:	34.5%	Conservative:	72		
Best Local Similarity:	23.2%	Mismatches:	241		
Query Match:	4.6%	Indels:	180		
DB:	3	Gaps:	34		
US-10-635-977-2 (1-541) x US-09-252-991A-12848 (1-3027)					
Qy	16	SerArgSerArgSerLysProArgAspGlnArgGluGluAlaGlySerSerAspLeu	35		
Db	1258	GCTCGATCCGCGCCCTCGCCCGGACGCGCGAAGAGCAACTGGTGCTGCG---	1311		
Qy	36	SerSerArgGlnAspAlaGluAsnAlaGluAlaLysLeuArgGly-LeuPro---	54		
Db	1312	-----CTGGAACTCAACCCCGCGCTGCGCGTGGCGCTTCTCGAGCTTCTAC	135		
Qy	54	nLeuValAspIleAlaCysLysValCysGlnAlaTyLeuGlyGlnLeuGluHisGlu---	73		
Db	1357	CCTGAAGCCGGTGTGGCGCGCTTCCGAGCTTCTACTGTGGCGACCCCGAACCGCAGCG	141		
Qy	74	-----AspIleAspThrSerAlaAspAlaValGluAspLeuTh	86		
Db	1417	CAAGCTGGCGCTGGCGGCACCGGACGCGCGCGCTGCGCGAT---CTCGACGAGTTGCC	147		
Qy	86	rGluAlaGluTrpGluAspLeuThrGlnGlnTy+SerLeuValHisGlyAspAlaPh	106		
Db	1474	GCTGGCC---TGGGAAGGC-----GGCGAC-----	1491		
Qy	106	elleSerAsnSerArgAsnTy+PheSerGlnCysGlnAlaLeuLeuAsnArgIleTh+Se	126		
Db	1496	-----AACCTGGCGAGCTTCGCCGCCGCCCTACGCGAGTTTCAGGCTCGT---	154		
Qy	126	rValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIleIysProAlaAl	146		
Db	1541	-----CCGGCGACCCCGCGCAAGCGCTGCGC-----GC	1561		
Qy	146	lysSerArgGlyArgAspIleValCysMetAspArgValGluGluIleLeuGluLeu---	165		
Db	1570	CGAACTGGCGCCCTACAGCTCGAAGCGCTTGAGCTGGATCGAGACCTTCGCGCAACTCGA	1625		
Qy	166	-----AlaAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValGl	181		

QY 286 -ProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAspHisValGluPr 305
Db 235 CCGCGGGCCGCTCTGCTGACAGCGG---CGCTGATCATCGCGAGGTGGAATA 291
QY 305 oArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArg----- 321
Db 292 CTGTGAAGAGCGAGTTACCGAGTACGGCAGCGAC---GCCGCGGCATATCGGCATCTT 348
QY 322 -----AspPheArgProTrpLeuIleGluLeuAsnSerSerPr 334
Db 349 CGCCAATACCAACCGCGGTACCGAATCTCTCGGAATCCTC-----GCCGG 396
QY 334 oThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspTh 354
Db 397 CTTCTCCGCGCGTCTGTCGCGGGTCCACCTGCGACCTGCAGGAACGCTCAGCGCGGACAT 456
QY 354 rIleLysValAlaValAspArgSerCysAspIleGly-AsnPheGluLeuLeuTyrArg 374
Db 457 TGTCCGCGCGTCTGTCGAGCGGTCCACCGACCTCGGCATCATCGCGCGCGGTGGCGC 516
QY 374 lnProVal-ValGluPro-----ProProPheSerGlySerAspLeuCys 388
Db 517 CACCGCGCTGCAGATCTCATTTTCAGCACCGCGCGCTGTGTGATGGCGT----- 567
QY 389 ValAlaGlyValSerValArg-----AlaArgGlnValLeu 402
Db 568 GCCCGCGGCCATCTCGCTGCGCGGACAGCGGATACCGCTACGCGACACCTGCAGTT 627
QY 403 ProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArg 422
Db 628 CCAGCA-----TATCGCGCTCCCGCGGCGC 654
QY 423 GlyProSerAlaMetProAspProAlaGlnGlyProProSer----- 436
Db 655 CACCTTGACGCGCTTCTGTAGACGCGGTGAGCAGATGGCGAGCAGCTATCGTGGC 714
QY 437 ---ProAlaLeuGln-----ArgAspLeuGlyLeuLysGluGlyLys 449
Db 715 CATCGAGTCTCCAGCTTCGAGCGCATCTGCGGATGATCGAGGCGCGCTCGGATCGG 774
QY 450 GlyLeuProLeuAlaLeuAlaProLeuArgGlyAlaAlaGluSerGlyGly----- 467
Db 775 CTGTATCCGCGAGTCCGCGCGCGCCGACAGCGGAGATGAAGCTGGCGACCATCGA 834
QY 468 -----AlaAlaGlnProThrArgThrLysAlaAlaGlyLys 479
Db 835 ACTCGACGAGCCTCGCGGATCCGCGAGCGCAGCTGCTGTGTCGCGCTGGAAGCACT 894
QY 480 ValGluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyVal 499
Db 895 GCCGAGTTGCTCGCG-----GGCGTTGATCGAGGAGTTGCAGAGGTGGG--- 939
QY 500 ProValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGlu 519
Db 940 -----AGACACCTGTAGGCGGATACGCCCTCGCGTGTGTCGCG----- 981
QY 520 ProValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGly 538
Db 982 -----GGGCTCGGAGATCGCGGA-----CAACCCCGCAGCGGT 1017

RESULT 15

US-09-252-991A-12970/c
; Sequence 12970, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12970
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12970

Alignment Scores:
Pred. No.: 0.032 Length: 1251
Score: 129.50 Matches: 94
Percent Similarity: 33.9% Conservative: 42
Best Local Similarity: 23.4% Mismatches: 141
Query Match: 4.5% Indels: 124
DB: 3 Gaps: 22
US-10-635-977-2 (1-541) x US-09-252-991A-12970 (1-1251)
QY 201 PheLeuValThrAspTrpAsnPro-----LeuThrIleTrpPheTyrLysGluSer 217
Db 1236 TTTGCAATTACCAATGGCGGTTCGGATTTCCCGAACGCCATGG-----AGT 1189
QY 218 TyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeu 237
Db 1188 GCCCATGCATTTTCGATCTCGCGA-----CCTGGCGCTGTTCATCCA 1147
QY 238 CysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuPro 257
Db 1146 TATCCCGAGTCGCCAGCCT-----GACCAGGCGCGCGCTTCCT 1099
QY 258 AlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGly 277
Db 1098 -----TTCCCGCGCGCGCCAGCGCGGTCAAGC 1066
QY 278 Ala-----ValTrp-GlySerValIleTyr----- 285
Db 1065 GCTCGAAGCGCAGCTTGGCAGCGCTGTGTACCGCAGCAGCGCGGTGAACTGAC 1006
QY 286 -ProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAspHisValGluPr 305
Db 1005 CCGCGCGCGCGCTGCTGTACAGCAGCG---CGGCTGATCATCGCGAGGTGGAATA 949
QY 305 oArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArg----- 321
Db 948 CTGTGAAGAGCGAGTTACCGAGTACGGCAGCGAC---GCCGCGCGGCATATCCGATCTT 892
QY 322 -----AspPheArgProTrpLeuIleGluLeuAsnSerSerPr 334
Db 891 CGCCAATACCAACCGCGGTACCGAATCTCTCGGAATCCTC-----GCCGG 844
QY 334 oThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspTh 354
Db 843 CTTCTCCGCGCTCGTCCGCGGTTCACCGTACGACCTGCAGGAACGCTCAGCGCGGACAT 784
QY 354 rIleLysValAlaValAspArgSerCysAspIleGly-AsnPheGluLeuLeuTyrArg 374
Db 783 TGTCCGCGCGTGTGAGCGCTCCACGACCTCGGCATCATCGCGCGCGGTGGCGC 724
QY 374 lnProVal-ValGluPro-----ProProPheSerGlySerAspLeuCys 388
Db 723 CACCGCGCTGCAGATCTGTGATTTTCAGCACCGCGCTGTGTGATGGCGT----- 673
QY 389 ValAlaGlyValSerValArg-----AlaArgGlnValLeu 402
Db 672 GCCCGAGGCGCATCCGCTGCGCGGACAGCGGAGTACGCTCAGCGCACCTCGAGTT 613
QY 403 ProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArg 422
Db 612 CCAGCA-----TATCGCGCTCCCGAGGCGC 586
QY 423 GlyProSerAlaMetProAspProAlaGlnGlyProProSer----- 436

```

Db      585  CACCTGCGCGCTTCTGTACGAACGGGTGGAGCAGATGGGCGACAGCATATCGCTGCG 526
Qy      437  ---ProAlaLeuGln-----ArgAspLeuGlyLeuLysGluGlyLys 449
Db      525  CATCCAGGTCTCCAGCTTCGAGCGCATCTGCCGATGATCGAGGGCGGTTCGGCATCGG 466
Qy      450  GlyLeuProLeuAlaLeuAlaProLeuArgGlyAlaAlaGluSerGlyGly----- 467
Db      465  CGTGATCCCGCAGTCCCGCGCGGCCACAGCCGACGATGAAGCTGGCGACCATCGA 406
Qy      468  -----AlaAlaGlnProThrArgThrLysAlaAlaGlyLys 479
Db      405  ACTCGAGAGCCCTGGCGATCCGCGAGCGCCTCTGTTACGCGACCTGGAAGCACT 346
Qy      480  ValGluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyVal 499
Db      345  GCCGAGTTGCTGCG-----GGCGTTGATCGAGGAGTTGCAGAGGTGGG--- 301
Qy      500  ProValAlaGlnProAlaLysSerTrpAppProAsnGlnLeuAsnAlaHisProLeuGlu 519
Db      300  -----AGACAGCCTGTAGGGCGGATACGCCCTGGCGTTGTCCGCG----- 259
Qy      520  ProValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGly 538
Db      258  -----GGGCTCCGGAGATCGGCGGA-----CAACCGCCAGGCGGT 223

```

Search completed: April 4, 2006, 09:44:02
 Job time : 333 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2006, 09:35:51 ; Search time 5994 Seconds
(without alignments)
4222.858 Million cell updates/sec

Title: US-10-635-977-2

Perfect score: 2854

Sequence: 1 MASSILKVVVSHQSCSRSR.....LRGLKTAEGALRPPGKGKS 541

Scoring table:

	BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0			0.5
Ygapop 6.0			7.0
Delop 6.0			7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```
-MODEL=frame+p2n.model -DEV=xlh
-Q=/abes/ABSSWEB/spool/US10635977/runat_04042006_103532_10020/app_query.fasta_1
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCAALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abes07
-USER=US10635977 @CGN 1 1 5315 @runat_04042006_103532_10020 -NCPU=6 -ICPU=3
-NO MAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database :

```
EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hic:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_ges1:*
10: gb_ges2:*
11: gb_ges3:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2129.5	74.6	1259	10	AY415398	AY415398 Homo sapi
2	1963.5	68.8	1257	10	AY415399	AY415399 Pan trogl
3	1755	61.5	2969	4	AK030151	AK030151 Mus muscu
4	1732.5	60.7	2720	4	AK029745	AK029745 Mus muscu
5	1639	57.4	1265	10	AY415400	AY415400 Mus muscu
6	1362	47.7	770	7	CR745100	CR745100 CR745100
7	1052	36.9	2520	4	AK080321	AK080321 Mus muscu

8	1020	35.7	4184	4	HS0808322	HS0808322
9	998	35.0	775	5	BU937372	BU937372 AGENCOURT
10	908	31.8	1059	11	DQ045868	DQ045868 Homo sapi
11	891	31.2	1023	3	BM808637	BM808637 AGENCOURT
12	811	28.4	1022	3	BM808516	BM808516 AGENCOURT
13	811	28.4	1059	11	DQ045869	DQ045869 Pan trogl
14	792.5	27.8	777	7	CK597795	CK597795 AGENCOURT
15	769	26.9	680	7	CJ018705	CJ018705 CJ018705
16	766.5	26.9	864	8	CD515431	CD515431 AGENCOURT
17	764.5	26.8	896	6	CD515431	CD515431 AGENCOURT
18	754	26.4	690	8	CD205198	CD205198 MNS08466
19	751	26.3	813	8	CD205198	CD205198 MNS08466
20	737.5	25.8	3328	4	BC028169	BC028169 Homo sapi
21	731	25.6	693	3	BM717061	BM717061 UI-E-EJ0-
22	701	24.6	679	7	CR986757	CR986757 CR986757
23	700	24.5	553	8	DN831611	DN831611 MGC1.7.1.
24	698	24.5	762	6	CD627148	CD627148 56047432J
25	694	24.3	776	7	CR998661	CR998661 CR998661
26	673	23.6	1371	8	DN726875	DN726875 CNB32-G05
27	671	23.5	861	7	CK653583	CK653583 AGENCOURT
28	659.5	23.1	787	8	DR722973	DR722973 AGENCOURT
29	657	23.0	503	7	CK604006	CK604006 AGENCOURT
30	644.5	22.6	616	7	CN349424	CN349424 170006000
31	639.5	22.4	616	5	BM740389	BM740389 BM740389
32	638	22.4	651	7	CF999072	CF999072 AGENCOURT
33	604	21.2	812	6	CD648578	CD648578 AUF_103_C
34	588	20.6	680	6	CB854808	CB854808 UI-CF-EC1
35	587	20.6	583	7	CN349423	CN349423 170004243
36	577	20.2	541	8	DN831612	DN831612 MGC1.7.1.
37	560	19.6	514	7	CR746623	CR746623 CR746623
38	520	18.2	831	8	DN532464	DN532464 1360510_M
39	513	18.0	320	1	AA144756	AA144756 mt73d04.r
40	490	17.2	724	2	BG916644	BG916644 602814292
41	485	17.0	285	8	DR008134	DR008134 MGC5.1.1.
42	476	16.7	880	5	BA448465	BA448465 BX448465
43	475	16.6	429	8	W31827	W31827 zb94f04.r1
44	474	16.6	504	3	BQ327883	BQ327883 MK4-R7004
45	470	16.5	580	3	BP305902	BP305902 BP305902

ALIGNMENTS

RESULT 1

AY415398

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

AY415398 Homo sapiens HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.
AY415398
AY415398.1 GI:39771357
GSS.
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1259)
Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarawal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios
Science 302 (5652), 1960-1963 (2003)
14671302
2 (bases 1 to 1259)
Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarawal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering


```

REFERENCE 2 (bases 1 to 1257)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
        Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
        Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
        Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
        Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
        them based on alignment.
FEATURES
    source      Location/Qualifiers
                1..1257
                /organism="Pan troglodytes"
                /mol_type="genomic DNA"
                /db_xref="taxon:9598"
                <!.>1257
                /locus_tag="HCM5544"
    gene
    ORIGIN
Alignment Scores:
Pred. No.:      1,11e-160      Length:      1257
Score:          1963.50      Matches:    386
Percent Similarity: 77.6%      Conservative: 2
Best Local Similarity: 77.2%      Mismatches: 25
Query Match:      68.8%      Indels:     86
DB:              10          Gaps:        1
US-10-635-977-2 (1-541) x AY415399 (1-1257)
QY 1 MetAlaSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerArg 20
DB 18 ATGGCATCCAGATCCTCAAGTGGGTGTGTCAGCCACAGAGCTGCAGAGGCGGCAGA 77
QY 21 SerLysProArgAspGlnArgGluAlaGlySerSerAspLeuSerSerArgGlnAsp 40
DB 78 AGCAAGCCAGGACACAGGAGGAGGCGGGAGCAGCGACCTGAGCAGGACGACGACAGAT 137
QY 41 AlaGluAenAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys 60
DB 138 GCTGAATATGTGTAGGCAAAAGCTCAGGGGCTCCCGGGGCGAGCTTGTGGACGTGCGTGC 197
QY 61 LysValCysGlnAlaTyrluGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp 80
DB 198 AAGGTGTCCAGGCGCTACTCGGGCAGCTGGAGCATGAGGACATGCACACGTCAGCAGAT 257
QY 81 AlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrTrpSerLeu 100
DB 258 GCCGTGGAGGACCTCAGTGGCCGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC 317
QY 101 ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu 120
DB 318 GTTCANNNNNATGCTTTTCATCTCCAATTCAGAAATTAATTTTCGNANNCCAGGCTCTG 377
QY 121 LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrp 140
DB 378 CTGAANNNNNTCCANNNTNTNNNCNTNAGACGGACATTTGACGGGNNCCGGAACATCTGG 437
QY 141 IleIleLysProAlaAlaLysSerArgGluValArgAspIleValCysMetAspArgValGlu 160
DB 438 ATTATAAGCCCGCGGCGCAAGTCCCGGGCGCGAGACATAGTGTGATGACCGCGGTGGAG 497
QY 161 GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValVal 180
DB 498 GAGATCCTGGAGCTGGCAGCTGCAGACCACTCTTTCCAGGGACACAACTGGGTGTC 557
QY 181 GlnLysTyrlleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrp 200
DB 558 CAGAAGTACATCGAGACCGCGCTCATCTATGACACCAAGTTCGACATCAGACAGTGG 617
QY 201 PheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArg 220
DB 618 TTCCTCGTCAGGAGTGAACCCCTGACCATCTGGTCTTCAAGGAGAGGTACTTTCGGG 677

```

```

QY 221 PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn 240
DB 678 TTCTCAACACACGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGTCAACAC 737
QY 241 AlaValGlnLysTyrluLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn 260
DB 738 GCCGTCCAGAAGACCTGAAGAATATGTGGGCCGACGCCCTCTGTGCCCGCACACAC 797
QY 261 MetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlnArgGlyAlaValTrp 280
DB 798 ATGTGACACGCGCCAGGTTCCAGGAGTACCTGCGAGCGCCAGGCGCGCGCTGTGG 857
QY 281 GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln 300
DB 858 GGCNNNGTCATCTACCCGTCCATGAAGAGGCGCATCGCCACACCATGAAGTGGCCAG 917
QY 301 AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly 320
DB 918 GACCACGTGGAGCCTTCGCAAGACAGCTTGAAGCTCTACGGGGCTGACTTCGTCTCTGGG 977
QY 321 ArgAspPheArgProTrpLeuIleGluLeuLeuLeuTrpArgGlnProValValGluPro 340
DB 978 AGGACCTTCAGNCCCTGGCTGATCGAGATCAATTCACGCCCCACCATGCACCCGTCACG 1037
QY 341 ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp 360
DB 1038 CCGGTACGCGCCAGCTGTGTGCACAGGTGCAGGAGGACACCATCAAGGTGCGCGTGAC 1097
QY 361 ArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValValGluPro 380
DB 1098 CGCAGCTGTGACATCGGCAACTTCGAGNNCCTGTGGAGGCGNNNNNGTGTGTGAGCGGCC 1157
QY 381 ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaArgGln 400
DB 1158 CCATTACGCGGTGCGACCTCTGTGGCGGGCGTCTAGTAG----- 1198
QY 401 ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys 420
DB 1198 ----- 1198
QY 421 AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln 440
DB 1198 ----- 1198
QY 441 ArgAspLeuGlyLeuLysGluGluLysGlyLeuProLeuAlaLeuAlaProLeuArg 460
DB 1198 ----- 1198
QY 461 GlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal 480
DB 1198 ----- 1198
QY 481 GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro 500
DB 1199 ---CTCCCGCGCTCCCTGTGCGCAGCTGGACAGTGCAGGCCCAACACCGCGTGTCCCC 1255
RESULT 3
AK030151 2969 bp mRNA linear HTC 03-APR-2004
LOCUS Mus musculus adult male testis cDNA, RIKEN full-length enriched
DEFINITION library, clone:4933401B17 product:hypothetical Tubulin-tyrosine
        ligase containing protein, full insert sequence.
ACCESSION AK030151
VERSION AK030151.1 GI:26326134
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
        Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
        Sciurognathi; Muridae; Murinae; Mus.
REFERENCE Carninci,P. and Hayashizaki,Y.
AUTHORS High-efficiency full-length cDNA cloning
TITLE

```

JOURNAL PUBMED REFERENCE AUTHORS	Meth. Enzymol. 303, 19-44 (1999) 10349636
TITLE	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 11042159
JOURNAL PUBMED REFERENCE AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Katsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,M., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,K., Izawa,M., Tanaka,T., Matsura,S., Kawai,J., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 11076861
TITLE	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
JOURNAL PUBMED REFERENCE AUTHORS	Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
TITLE	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
JOURNAL PUBMED REFERENCE AUTHORS	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 6 (bases 1 to 2969)
TITLE	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toyota,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.jp/ URL:http://fantom.gsc.riken.jp/.
FEATURES source	Location/Qualifiers 1. .2969 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM,DB:4933401B17" /db_xref="taxon:10090" /clone="4933401B17" /sex="male" /tissue_type="testis" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult"

QY 200 TrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeu 219
 DB 1696 TGGTTCTCTGCACAGACTGGAAATCCCTTAACCATCTGGTTCTACAAAGAAAGCTACCTC 1755
 QY 220 ArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCyAsn 239
 DB 1756 CGCTTCTCACACAGGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGGTGAAC 1815
 QY 240 AsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHis 259
 DB 1816 AACTCCATCCAGAGCGCTCAAAAATGATAAGAGCGCAGTCCGCTGCTACCTTGTCTAT 1875
 QY 260 AsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaVal 279
 DB 1876 AACATGTGGACAGCACTCGTTTCCAGGAGTAGTCTGCAGAGAGGCGCGAGGAGGAACG 1935
 QY 280 TrpGlySerValIleTyrProSerMetLysValAlaIleAlaHisAlaMetLysValAla 299
 DB 1936 TGGGTAGCATCACTACCCGCTATGAAGAGCTGTCAACCAATGCCATGAGGGTAGCC 1995
 QY 300 GlnAspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeu 319
 DB 1996 CAGGACCACTAGAACCGGTAAGAACAGCTTCGAGCTCTACGGAGCGGACTTATCATCTG 2055
 QY 320 GlyArgAspPheArgProTrpLeuIleGluLysAsnSerSerProThrMetHisProSer 339
 DB 2056 GGGCGAGACTTCAAGCCATCGCTTATCGAGATCAACTCCAGCCCAACCATGCCATGCCCTCC 2115
 QY 340 ThrProValThrAlaGlnLeuCyAsnAlaGlnValGlnGluAspThrIleLysValAlaVal 359
 DB 2116 ACTCCCGTCACAGCTCAGCTCTGTGCCAGGTGAGGAGGACACCATCAAGGTGGTGTG 2175
 QY 360 -----AspArgSerCysAspIleGlyAsnPheGluLeuLeuTyrArgGlnPro 375
 DB 2176 GATCGAACTGGACCGAACTGTGACATCGCACTTTGAGCTTCTGTGGCGGAGCCT 2235
 QY 376 ValValGluProProProPheSerGlySerAspLeuCyValAlaGlyValSerValArg 395
 DB 2236 GCCGTGGAGTGCACCGCTTCAACGGGTCTGACCTCTGTGGAGGATATCAGCGTGAAG 2295
 QY 396 ArgAlaArgArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAsp 415
 DB 2296 AAAGCCAGAACAAATGCCCGCCATTCAGCGCTCGGCCCTCTCAGAATCACTCTTAGAT 2355
 QY 416 AlaGlnProLeuLysAlaArgLysProSerAlaMetProAspProAlaGlnGlyProPro 435
 DB 2356 GCTCCACCC---AAAGTCGGAGTGCCTGGCGGCTGATGGAAACAGTATCAGGCCCTCCA 2412
 QY 436 SerProAlaLeuGlnArgAspLeuGlyLysGluLysGlyLeuProLeuAlaLeu 455
 DB 2413 AGAACGACCGTGGCGGAGGAGTGGAAACGGGAAGCAAGGTACTTTCAACCACTGG 2472
 QY 456 LeuAlaProLeuArgGlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLys 475
 DB 2473 TCCATGCCAGTAATGGAT---GCAGAGGTTCAGAGGTTCAGCAAGGCCATCTATGCATTC 2529
 QY 476 AlaAlaGlyLysValGluLeuProAlaCysProCysArgHisValAspSerGlnAlaPro 495
 DB 2530 GAGTCAATGACTACCA-----CATGTAGCAATAAAGTCCAC 2568
 QY 496 AsnThrGly-ValProValAlaGlnProAlaLysSerTrpAsp---ProAsnGlnLeuAs 514
 DB 2569 AAGAGCGGTATACAGAGTCCAAATCCAGTAAGTCCAGGGGTAACTTTAAACAAGTCA 2628
 QY 514 nAlaHisProLeuGluProValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgPr 534
 DB 2629 CAG-CACCCCT---GCCCTGTTTGCACAGACTATGAAACCCACACAGATGACCTCGTCTCC 2684
 QY 534 oProGlyGlyLysGlySer 541
 DB 2685 ACCACCCACTGCTGAGGTAAAC 2706

RESULT 4

AK029745

LOCUS

DEFINITION

Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930524K07 product:hypothetical tubulin-tyrosine ligase containing protein, full insert sequence.

ACCESSION

AK029745

VERSION

AK029745.1

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

REFERENCE

Carninci, P. and Hayashizaki, Y.

TITLE

High-efficiency full-length cDNA cloning

JOURNAL

Meth. Enzymol. 303, 19-44 (1999)

PUBMED

10349636

AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL

Genome Res. 10 (10), 1617-1630 (2000)

REFERENCE

11042159

AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

TITLE

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL

Genome Res. 10 (11), 1757-1771 (2000)

REFERENCE

11076861

AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.

TITLE

Functional annotation of a full-length mouse cDNA collection

JOURNAL

Nature 409, 685-690 (2001)

REFERENCE

5

AUTHORS

The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL

Nature 420, 563-573 (2002)

REFERENCE

6

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE

Direct Submission

JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.jp/>

URL: <http://fantom.gsc.riken.jp/>

FEATURES

source

```
1. .2720
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:4930524K07"
/db_xref="taxon:10090"
/clone="4930524K07"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
934. .2430
```

CDS

```
/note="unnamed protein product; hypothetical
Tubulin-tyrosine ligase containing protein
(InterPro:IPR004344, evidence: InterPro)
putative"
/codon_start=1
/protein_id="BAC26595.1"
/db_xref="GI:2632582"
```

```
/translation="MLNHYGKTSFTTKIGLCLNMRSLPWYVOANPTFFPCYGLCT
ESEKQFLDDFRTRVAASILKWLVLHQNYSKVKSEAKNSDPSPKKDPENPDL
KLPSLGSQVVDACKVQVAYLGOLEDHEDIVSEASTEALSEEWNDLTOQYLLVHGN
ASTDSKSPFAOCCALWKSISVNPOTEDIGRNWIIPAAKSRGRDIVCMRDVENI
LSLVAADSTQTKDNKWKVQYLETPLMIYDTKFDIRQWFLVTDNPLTIWFYKESYLR
FSTQFSLKLDGSAIHLNNSIQRLNKNKERSPLLPCHNMTSIFQSYLQKRGEGG
TWGSIITPYMKRAVTNMAVQDHVEARKNSFELYGADFILGRDFPKPLIELINSSPTM
HPSTPTAOLCAQVEDTIKVVDVRKLDNRCDIGNFELLWRQPAVSLPFPNGSDLCVE
GISVKKAKQMPPIASVGLSESLLDAPPKVRARALMETVIRPPRTTVRQDKREAK
GFRVISKRRRA"
```

ORIGIN

Alignment Scores:

```
Pred. No.: 5,3e-140 Length: 2720
Score: 1732.50 Matches: 337
Percent Similarity: 80.7% Conservative: 60
Best Local Similarity: 68.5% Mismatches: 81
Query Match: 60.7% Indels: 15
DB: 4 Gaps: 5
```

US-10-635-977-2 (1-541) x AK029745 (1-2720)

```
Qy 1 MetAlaSerSerIleLeuLysTyrValValSerHisGlnSer---CysSerArgSerSer 19
Db 1108 GTGGCTGCCAGCATCTCCAAAGTGGGTAGTCTTCCACCAAGAAATTAAGTCC-----AGC 1158
Qy 20 ArgSerLysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerSerArgGln 39
Db 1159 AAGTCAAGGCGCAGAGTAGAGAGGAGGAGGCGCCAGAACAGTACGCCACCCCAAGAAA 1218
Qy 40 AspAlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlnGlnLeuValAspIleAla 59
Db 1219 GATCCTCAAAATCTCGACCTCAAGCTACCAAGCCCTCTCGGACACAGGTGTAGACACAGCT 1278
Qy 60 CysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAla 79
Db 1279 TGCAAGGTGTGCCAGGCTTACCTCGGACAGCTAGACATGAAGACATAGACGTATCGGAG 1338
Qy 80 AspAlaValGluAspLeuThrGluAlaGluTyrGluAspLeuThrGlnGlnTyrTyrSer 99
Db 1339 GCCAGCACCGAGGCTTGTCTGAGAGGAATGGAAATGACCTGACACAGCAGTACTATCTA 1398
Qy 100 LeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAla 119
Db 1399 CTGTTTCATGGCAATGCTTCCATCACTGATTCGAAAAGTTACTTTGGCGCAATGCCAAGCT 1458
Qy 120 LeuLeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIle 139
Db 1458 TCCCTGCGCCTTGATGTCGCGAGGCGCCTTCTGCTG 2570
```

```
Db 1459 CTGCTGAGTAAGATCAGTTCTGTGAACCTTCAGACAGAGATCGATGGGATCCGGAACATC 1518
Qy 140 TrpIleIleLysProAlaLysSerArgGlyArgAspIleValCysMetAspArgVal 159
Db 1519 TGGATCATAAAGCCTGCAGCCAAAGTCCCGGGTTCGAGATATTGTGTATGACCGGTG 1578
Qy 160 GluGluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTyrVal 179
Db 1579 GAGAATCTCTGAGTCTGGTGGCCGACAGACAGCCAGACCAAGGACCAACAAATGGGTG 1638
Qy 180 ValGlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGln 199
Db 1639 GTCCAGAGTACATCGAGACCCCAATGCTCATCTACGACACCAAGTTTGACATCAGACAG 1698
Qy 200 TrpPheLeuValThrAspTyrAsnProLeuThrIleTyrPheTyrLysGluSerTyrLeu 219
Db 1699 TGGTTCCTCGTCACAGACTGGAAATCCCTTAAACATCTGGTCTTACAAAAGAAAGTACCTC 1758
Qy 220 ArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsn 239
Db 1759 CGCTTCTCCACACAGCGCTTCTCCCTGGACAGCTGGACAGCGCCATCCACCTGTGTAAAC 1818
Qy 240 AsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHis 259
Db 1819 AACTCCATCCAGAGCGCCTCAAAATGATAAAGAGCGCAGTCCGCTGTCTTCTCAT 1878
Qy 260 AsnMetTyrThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyValArgGlyAlaVal 279
Db 1879 AACATGTGACAGCAGTCTGTTCCAGGAGTACCTCGAAGAGAGGCGCCGAGGAGAACG 1938
Qy 280 TrpGlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAla 299
Db 1939 TGGGGTAGCATCATCTACCGCTATGAAAGAGCTGTCCACCAATGCCATGAGGGTAGCC 1998
Qy 300 GlnAspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeu 319
Db 1999 CAGGACCACTAGAGCCGCTAAGAACAGCTTCGAGCTTCTACGAGCGCAGCTTCATCCTG 2058
Qy 320 GlyArgAspPheArgProTyrLeuIleGluLeuAsnSerSerProThrMetHisProSer 339
Db 2059 GGGCGAGACTTCAAGCCATGGCTTATCGAGATCAATCTCAGCCCCACCATGACCCCTCC 2118
Qy 340 ThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaVal 359
Db 2119 ACTCCGTCACAGCTCAGCTCTGTGCCAGGTGTCAGAGGACACCATCAAGGTGGTGTG 2178
Qy 360 -----AspArgSerCysAspIleGlyAsnPheGluLeuLeuTyrArgGlnPro 375
Db 2179 GATCGCAAACTGGACCGAACTGTGACATCGGCAACTTTGAGCTTCTGTGGCGCAGCCT 2238
Qy 376 ValValGluProProPheSerGlySerAspLeuCysValAlaGlyValSerValArg 395
Db 2239 GCGTGTGAGCTGCCCGTTCAACGGGTCTGACCTCTGTGTGAAGGTATCAGCGTGAAG 2298
Qy 396 ArgAlaArgArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAsp 415
Db 2299 AAAGCCAGAGCAATGCGCCCATGTGCCAGCGTCCGCTCTCAGAAATCACTCTTAGAT 2358
Qy 416 AlaGlnProLeuLysAlaArgGlyProSerAlaMetProAspProAlaGlnGlyProPro 435
Db 2359 GCTCCACCC---AAAGTGGAGTGGCCGCGCTGATGGAAACAGTGTATCAGGCGCTCCA 2415
Qy 436 SerProAlaLeuGlnArgAspLeuGlyLeuGluGluLysGlyLeuProLeuAlaLeu 455
Db 2416 AGAACGACCGTGGCGGAGGACTGGAAACGGGAAGAGC-AAAGGGCTTTCGTGTGATCTC 2474
Qy 456 LeuAlaProLeuArgGlyAlaAla-----GluSerGlyGlyAlaAlaGln 470
Db 2475 CAAGCGCGCTGCTAGTTCTGTGGTCTCCGCCCGCCCGCTGCGCTGCGCAAGCTC 2534
Qy 471 ProThrArgThrLysAlaAlaGlyLysValGluLeu 482
Db 2535 TCCCTGCGCCTTGATGTCGCGAGGCGCCTTCTGCTG 2570
```

```

RESULT 5
AY415400      1265 bp      DNA      linear      GSS 12-DEC-2003
LOCUS      Mus musculus HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION      genomic survey sequence.
ACCESSION      AY415400
VERSION      AY415400.1 GI:39771359
KEYWORDS      GSS.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
1 (bases 1 to 1265)
Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
Todd, M.A., Tanenbaum, D.M., Civeello, D.R., Lu, F., Murphy, B.,
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE      Inferring nonneutral evolution from human-chimp-mouse orthologous
JOURNAL      gene trios
PUBMED      Science 302 (5652), 1960-1963 (2003)
14671302
2 (bases 1 to 1265)
Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
Todd, M.A., Tanenbaum, D.M., Civeello, D.R., Lu, F., Murphy, B.,
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE      Direct Submission
JOURNAL      Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT      This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES      Location/Qualifiers
source      1..1265
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            <1..>1265
            /locus_tag="HCM5544"
gene
ORIGIN
Alignment Scores:
Pred. No.:      2,53e-132      Length:      1265
Score:      1639.00      Matches:      305
Percent Similarity:      87.4%      Conservative:      43
Best Local Similarity:      76.6%      Mismatches:      42
Query Match:      57.4%      Indels:      8
DB:      10      Gaps:      3

US-10-635-977-2 (1-541) x AY415400 (1-1265)

QY      1      MetAlaSerIleLeuLysTrpValValSerHisGlnSer---CysSerArgSerSer 19
DB      18      GTGGCTGCAGCATCTCAAGTGGGTAGTCTTCCACCAAGTACTGCG-----AGC 68
QY      20      ArgSerLysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerSerArgGln 39
DB      69      AAAGTCAAGGGCAAGAGTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 128
QY      40      AspAlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAla 59
DB      129      GATCTGAAATTCCTGACCTCAAGCTTACCAAGCTCTCGGGACAGGTTGTAGACACAGCT 188
QY      60      CysLysValCysGlnAlaTyrluGlyGlnLeuGluHisGluAspIleAspThrSerAla 79
DB      189      TGCAGAGGTGTCCAGGCTTACCCTGGACAGCTAGACATGATGAGATGATGATGATGATG 248
QY      80      AspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnTrpTyrSer 99
DB      249      GCCAGCACCAGGCGCTTCTCTGAGGAGGAGGATGATGATGATGATGATGATGATGATG 308
QY      100      LeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAla 119

```

```

Db      309      CTGGTTTCATGGCAATGCTTCCATCATGATTGAAAAAGTTACTTTTCCGAATGCCAAGCT 368
QY      120      LeuLeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIle 139
Db      369      CTGCTGAGTAAGATCAGTCTGTGAACCTCAGACAGAGATCAGTGGGATCGGAACATC 428
QY      140      TrpIleIleLysProAlaLysSerArgGlyArgAspIleValCysMetAspArgVal 159
Db      429      TGGATCATAAAGCTTCGAGCAAGTCCCGGGTCCAGATATTGTGTGTATGACCGGTG 488
QY      160      GluGluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAsnLysTrpVal 179
Db      489      GAGAACATCTCTGAGTCTGTGTGCGCCAGACACCCAGACCAACAAAGACCAAAATGGGTG 548
QY      180      ValGlnLysTrpIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGln 199
Db      549      GTCCAGAGTATACGAGACCCCAATGTCTATCAGACACCCAGTTTGACATCAGACAG 608
QY      200      TrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeu 219
Db      609      TGGTTCTCTCGTCACAGACTGGAATCCCTTAACCATCTGTGTTCTACAAAGAAAGTACCTC 668
QY      220      ArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsn 239
Db      669      CGCTTCTCCACACAGCGCTTCTCCCTGGACAGCTGGACAGCGCCATCCACCTGTGTAAC 728
QY      240      AsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHis 259
Db      729      AACTCCATCCAGAGCGGCTCAAAATGATAAGAGCGAGTCCGCTGTCTACTTGTCTAT 788
QY      260      AsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaVal 279
Db      789      AACATGTGGACCCAGCACTCGTTTCCAGGAGTACCTTCAGAGAGGCGCGGAGGAGAACG 848
QY      280      TrpGlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAla 299
Db      849      TGGGGTAGCATCATCTACCCGCTTATGAAAAGAGCTGTCCACCAATGCCATGAGGGTAGCC 908
QY      300      GlnAspHisValGluProArgLysAsnSerPheGluLeuTyrGlyValAspPheValLeu 319
Db      909      CAGGACCACTAGAACGCCGTAAAGAACAGTTTCGAGCTCTACGAGACCCAGCTTCATCCTG 968
QY      320      GlyArgAspPheArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSer 339
Db      969      GGGCGAGACTTCAAGCCATGCTTATCGAGATCACTCCAGCCCCCAGCCATGACCCCTCC 1028
QY      340      ThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaVal 359
Db      1029      ACTCCCGTCACAGCTCAGCTCTGTGCCAGGTGCGAGGAGGACCAATCAAGGTGGTGTG 1088
QY      360      -----AspArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnPro 375
Db      1089      GATCGCAAACTGGACCGCAACTGTGACATCGCAACTTGTAGCTTCTGTGGCGGAGCCT 1148
QY      376      ValValGluProProPheSerGlySerAspLeuCysValAlaGlyValSer 393
Db      1149      GCCGTGGAGTGCACACCGTTCACCGGTCTGACCTCTGTGTGGAGGTATCAGC 1202

RESULT 6
CR745100      770 bp      mRNA      linear      EST 21-DEC-2004
LOCUS      CR745100 Soares testis NHT Homo sapiens cDNA clone IMAGE998M164174
DEFINITION      IMAGE:1645239-5, mRNA sequence.
ACCESSION      CR745100
VERSION      CR745100.1 GI:51667573
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE      1 (bases 1 to 770)

```

AUTHORS

Ebert, L., Heil, O., Hennig, S., Korn, B., Neubert, P., Partsch, E.,
Peters, M., Radelof, U. and Schneider, D.
I.M.A.G.E. cDNA Clone Collection
Unpublished (2004)

TITLE

JOURNAL

COMMENT

Contact: Inge Arlart

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Heubnerweg 6, D-14059 Berlin, Germany

Email: www.rzpd.de

RZPD; IMAG998M164174.

RZPDLIB; I.M.A.G.E. cDNA Clone Collection;

Contact: Inge Arlart

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Heubnerweg 6, D-14059 Berlin, Germany

Tel: +49 30 32639 100

Fax: +49 30 32639 111

www.rzpd.de

This clone is available royalty-free from RZPD;

Contact RZPD (clone@rzpd.de) for further information. Seq primer:

M13r, Primer sequence: TTTACACAGGAACAGCTATGAC.

FEATURES

source

1. .770
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAG998M164174 ; IMAGE:1645239"
/sex="male"
/lab_host="DH10B"
/clone_lib="Soares testis NHT"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5].
TGTTACCAATCTCAAGTGGAGCGCGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cos5, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Alignment Scores:

Pred. No.: 1.94e-108 Length: 770
Score: 1362.00 Matches: 254
Percent Similarity: 99.2% Conservative: 0
Best Local Similarity: 99.2% Mismatches: 2
Query Match: 47.7% Indels: 0
DB: 7 Gaps: 0

US-10-635-977-2 (1-541) x CR745100 (1-770)

Qy 85 LeuThrGluAlaGluTrpGluAspLeuThrGlnTrpTyrSerLeuValHisGlyAsp 104
Db 3 CTCACCTAGGCCGAGTGGAGGACCTGACCCAGCAGTACTACCTCGTTTCATGGCGAT 62
Qy 105 AlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeuLeuAsnArgIle 124
Db 63 GCCTTTCATCTCCAATTCACGAATTTACTTTTCGCAGTGCCAGGCTCTGCTGAATAGATC 122
Qy 125 ThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIleIleLysPro 144
Db 123 ACGTCTGTGAACCTTCAGACGACATTCAGCGGGTCCGGAACTCTGGATTATTAAGCCC 182
Qy 145 AlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGluLeuGlu 164
Db 183 GCGGCCAAGTCCCGGGCCGAGACATAGTGTGCATGCACCGTGTGGAGGAGATCCTGGAG 242
Qy 165 LeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValGlnLysTyrIle 184
Db 243 CTGGCAGCTGCAGACCAACCTCTTTTCAGGGACAAACAGTGGTGGTCCAGAGTACATC 302
Qy 185 GluThrProLeuIleCysAspThrLysPheAspIleArgGlnTrpPheLeuValThr 204

Db 303 GAGACGCCGCTGCTCATCTGTGCACACAAAGTTTCGACATCAGACAGTGGTTCCTCGTCACG 362
Qy 205 AspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPheSerThrGln 224
Db 363 GACTGGAAACCCCTGACCATCTGGTTCTACAAGGAGATTACTTGGCGTTCTCAACTCAG 422
Qy 225 ArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuLeuCysAsnAsnAlaValGlnLys 244
Db 423 CGCTTCTCCCTGGACAAAGCTGGACAGCGCATCCACTGTGCAACAACGCCGCTCCAGAG 482
Qy 245 TyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMetTrpThrSer 264
Db 483 TACCTGAAGAATGATGTGGCGCGCAGCCCTGCTGCCCGCACACAACATGTGGACACG 542
Qy 265 ThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyValAlaValTrpGlySerValIle 284
Db 543 ACCAGGTTCCAGAGTACCTGCAGCGCCAGGGCCGTGGCGCGCTGTGGGCGACGCTCATC 602
Qy 285 TyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAspHisValGlu 304
Db 603 TACCCGTCCATGAGAAAGCCCATCGCCACGCCATGAAGTGGCCACGACACGCTGGAG 662
Qy 305 ProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArgAspPheArg 324
Db 663 CCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCCTGGGANGGACTTCANG 722
Qy 325 ProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThr 340
Db 723 CCCTGGCTGATGAGATCAATTCACGCCCCACCATGACCCCGTCCACG 770
RESULT 7
AK080321 2520 bp mRNA linear HTC 03-APR-2004
LOCUS Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
DEFINITION library, clone:A630053H17 product:HOTTI, PROTEIN homolog [Homo
sapiens], full insert sequence.
ACCESSION AK080321.1 GI:26348476
VERSION HTC; CAP trapper.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
JOURNAL prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
JOURNAL sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)

REFERENCE
AUTHORS

5 The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

REFERENCE 6 (bases 1 to 2520)

AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Numata, K., Nomura, K., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

DIRECT SUBMISSION

TITLE Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.jp, URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

A cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/WRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge. Please visit our web site for further details.

URL: <http://genome.gsc.riken.jp/>
URL: <http://fantom.gsc.riken.jp/>

FEATURES
source

1. .2520
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:A630053H17"
/db_xref="taxon:10090"
/clone="A630053H17"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stages="3 days neonate"
84. .2198
/note="unnamed protein product; HOTT1L PROTEIN homolog [Homo sapiens] (SPT109199, evidence: FASTY, 85.8%ID, 93.4%length, match=821)
putative"

CDS

/codon_start=1
/db_xref="BAC37878.1"
/db_xref="GI:26348477"
/translation="WDTSSPFVAAGQGEASPPSLPVSTISPSLOSVMVNETPYLIWT
TERDVLDCRFLSKQMINHYARAGSTTKVGLCLNRLNLPWFEDADASFPFCVRLG
AEDDKATIEFWLTAARLVKLVVKKSEKQSISIQAREBEAEDTPQKQKKLV
VSSDFDALSACQHLSSIAHKDIDKPNPLFLVSDPDWSQFLQRYQIVHEGAELR
YLEVQVQCEDILOQLNVVQPDMDGDRNIWVPGAKSRGIMCMNRLDMLKLV
DCNPLMKDGMWVOKYIERPLLIPTGKFDLQWPLVTDWNLTVWFYRDSYIRFSTQ
PFLNKLNDNSVHLNCSIQRHLEASCRHPMLPDPNMWSSORFOAHLEQVDPKAWSS
VIVFGMKAIVHALQTSQDNQCRKASFEIYAGDFVFGEDFQPMLEINAFSTWAPST
AVTARKCAGVQADTLRVVDRRLDKSCDTGAFELIYKQPAVEVQYVQIRLLVEGSLI
KKPVVGHRRTCGVRSLLPHLLTQQGSKDSGSPFTHRSASRKNRAESLEHTEKPEP
AAVASVSGKGPAPFPPLSHKAWLPSPVRHPQGRVLRQLQDLQSKALSTTGA
LMTLPTAKVLSPFPDLKLAPSLMKPQKVGFEFCCTTWRVVLSSGIGERBGRQRAA
PRPSSAPCKGLSSTFPCSKTET"

ORIGIN

Alignment Scores: 1.37e-80 Length: 2520
Pred. No.: 1052.00 Matches: 226
Score: 1052.00
Percent Similarity: 55.1% Conservative: 85
Best Local Similarity: 40.1% Mismatches: 177
Query Match: 36.9% Indels: 76
DB: 4 Gaps: 8
US-10-635-977-2 (1-541) x AK080321 (1-2520)
QY 2 AlaSerSerIleLeuLysTrpValSerHisGlnSerCysSerArgSerArgSer 21
438 GCCCGCAAGCTCTCAAGCTGTGTGGTGAAGTGGAGAGAGAGTCAATCTCCATC 497
QY 22 LysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerArgGlnAspAla 41
498 CAGGCAAGAGAG-----GAAGAGGCCCCAGAGAGATACACAGCCCCAGAAACAG----- 545
QY 42 GluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLys 61
546 -----GAAAGAAAGCTAGTACCGGTCTCTCAGACTTTGTGATGAGGCTCTGAGT 596
QY 62 ValCysGlnAlaTyrluGlyGlnLeuGluHisGluAspIleAspThrSerAlaAspAla 81
597 GCATGCCAGGAGCCTTTAGCAGCATAGCCCCAAGACATTTGACAAAGACATCC 656
QY 82 ValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrluSerLeuVal 101
657 CGCTCTACCTCAGCCCGATGATGGTCCCAAGTTCCTTACGGCTTACCAATAGTT 716
QY 102 HisGlyAspAlaPheIleSerAsnSerArgAsnTyrluPheSerGlnCysGlnAlaLeu 121
717 CATGAAGGGGAGAGCTCAGATACCTGGAGGTCCAAAGTCCAGCGCTGTGAAGACATCTTA 776
QY 122 AsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIle 141
777 CAGCAGCTTCAGACGCTGGTACCCCAATGATGATGAGGGGGGATCGGAACATCTGGATC 836
QY 142 IleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluLys 161
837 GTGAAGCCCGGAGCCNAGTCCGAGCGGAGGATTTGTGATGATGAACCGCTGGATGAG 896
QY 162 IleLeuGluLeuAlaAlaAlaAspHisProLeuSerArgAspAsnLysTrpValGln 181
897 ATGCTGAAGCTGGTGGAGCTGCAACCCCATGCTCATGAAGGATGCGCAAGTGGATCGTCAG 956
QY 182 LysTyrlleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPhe 201
957 AAGTACATTGAGCGCGCCCTCTTATCTTCGGCACCAAGTTGACCTGACAGACAGTGGTTC 1016
QY 202 LeuValThrAspTrpAsnProLeuThrIleTrpPheTyrluLysGluSerTyrluArgPhe 221
1017 CTGTGACTGCTGGACCCACTACCGTGTGGTGTCTACCGAGAGAGCTACATTCGTTTC 1076
QY 222 SerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAla 241
1077 TCCACACAGCCCTTCTCCCTGAAGAACCTCGACAACTCTGTGCACCTGTGTAAACAATCC 1136
QY 242 ValGlnLysTyrluLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMet 261
1137 ATCCAGAGACACTTGGAGGCTTCTGTGTCACCGGACCCCGATGCTGCCCCAGATATCAATG 1196
QY 262 TrpThrSerThrArgPheGlnGluTrpLeuGlnArgGlnGlyArgGlyAlaValTrpGly 281
1197 TGTTCACAGAGAGTTCAGGCCCACTTTCAGGAGGTAGATGCCCAAGAGCCCTGGTCC 1256
QY 282 SerValIleTyrluProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAsp 301
1257 AGCGTCATCGTCCAGGATGAAGGCTGCTGTGATCCATCCCTGCACACCTCCCAAGAT 1316
QY 302 HisValGluProArgLysAsnSerPheGluLeuTyrluGlyAlaAspPheValLeuGlyArg 321
1317 AACGTGAGTCCCGAAGGCCAGCTTTGAGCTCTATGGGCGAGACTTTGTGTGGGAA 1376

Qy	322	AspPheArgProTrpLeuIleGluLeuAanSerSerProThrMetHisProSerThrPro	341	
Db	1377	GACTCCAGCCCTGGTGAATGAATGACAGCCGCCACCATGCGACTCCACGGCT	1436	
Qy	342	ValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaVal	359	
Db	1437	GTCACTGCCCGCTCTGTCCCGGTGTGACGAGATACCTGGCTGTGTATCATCGACCGG	1496	
Qy	360	-----AspArgSerCysAspIleGlyAanPheGluLeuLeuTrpArgGlnProVal	377	
Db	1497	CGCCTGAGCCGACGTGTGATACCGGAGCCTTTGAGCTCATCTATAAGCAGCCTGCTGTG	1556	
Qy	378	GluProProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAla	397	
Db	1557	GAGGTGCTCAGTACGTGGGTATCCGGCTCTAGTGGGAGGCTCTACCATCAAGAAG	1613	
Qy	398	ArgArgGlnValLeuProValCysAanLeuLysAlaSerAlaSerLeuLeuAspAlaGln	417	
Db	1613	-----	1613	
Qy	418	ProLeuLysAlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerPro	437	
Db	1614	-----CCCGTGCCC	1622	
Qy	438	AlaLeuGlnArgAspLeuGlyLeuLysGluGlyLysGlyLeuProLeuAlaLeuAla	457	
Db	1623	GTGGGTATCGGGGACAGGGGTCCGCTCATCA	1670	
Qy	458	ProLeuArgGlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAla	477	
Db	1671	ACCAGCAAGGCTCTGGGAAGCAAGGACTAGGTCCTCCCTACCCAC	1727	
Qy	478	GlyLysValGluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAanThr	497	
Db	1728	TCTAGGAATAATCTAGGGCCGAGAGCTGGAGCACACCGAGAGCGGAACCTGGGCC	1787	
Qy	498	GlyValProValAlaGlnPro	505	
Db	1788	GTGCGCTCTGCTCCGGAAGGGGAGAAAGCCCTTTCCACTTCCCTACGCTCCATCTC	1847	
Qy	506	LysSerTrpAspProAanGlnLeuAanAlaHisProLeuGluProValLeuArg	523	
Db	1848	AAGCCTGTGCTGCTCTCCCGGTGTGCACCGACCGCGCGGTCTCAGACTCCAA	1907	
Qy	524	-----GlyLeuLysThrAlaGluGlyAlaLeuArgProPro	535	
Db	1908	CATGACCAGCTGGTGGCTCTAAGGCTCTGTCAACACAGGCAAGGCTTTGATGACTCTT	1967	
Qy	536	ProGlyGlyLys	539	
Db	1968	CCTACTGCCAAG	1979	
RESULT 8				
HSM808322				
LOCUS				
DEFINITION Homo sapiens mRNA; cDNA DKFZp686D076 (from clone DKFZp686D076).				
ACCESSION BX648175				
VERSION BX648175.1 GI:34367334				
KEYWORDS HTC.				
SOURCE Homo sapiens (human)				
ORGANISM Homo sapiens				
REFERENCE				
AUTHORS				
1 (bases 1 to 4184)				
Ansoorge, W., Krueger, S., Regiert, T., Rittmueller, C., Schwager, B.,				
Mewes, H. W., Weil, B., Amid, C., Osanger, A., Fobo, G., Han, M. and				
Wiemann, S.				
The German cDNA Consortium				
Direct Submission				
Submitted (22-SEP-2004) MIPs, Ingolstaedter Landstr.1, D-85764				
Neuerberg, GERMANY				
CONSTRM				
TITLE				
JOURNAL				
COMMENT				
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by EMBL (European Molecular Biology Laboratories, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.				
This clone (DKFZp686D076) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany. Please contact RZPD for ordering:				
http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686D076				
Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.				
FEATURES				
Location/Qualifiers				
1..4184				
/organism="Homo sapiens"				
/mol_type="mRNA"				
/db_xref="RZPD:DKFZp686D076Q"				
/db_xref="taxon:9606"				
/clone="DKFZp686D076"				
/issue_type="cervix"				
/clone_lib="686 (synonym: hlcc3). Vector pSport1_sfi; host DH10B; sites SfiIA + SfiIB"				
/dev_stage="adult"				
/note="hypothetical protein"				
1..4184				
/gene="DKFZp686D076"				
2529..3587				
/gene="DKFZp686D076"				
/codon_start=1				
/product="hypothetical protein"				
/protein_id="CAH10554.1"				
/db_xref="GI:50949619"				
/translation="MDIDKDLAPLYLTPGWSLFLQRYVQVVEGAELRLDTQVQR				
CEDILQQLQAVVQPIQIDMEGRNITWIKPKGSKRGIMCMHLEMLKLVNGLNPVVMK				
DKGVVQVKTIERPLIFGKFDLRQWFLVDNPLTWFRDSYIRFSIQPFSLKILND				
NSVHLCNNSIQKHLENSCHRPDLPPDNWSSORFOAHOEGAPNAWSTIIVPGMKD				
AVTHALQTSQTVQCRKASFELYGADVFEGEDFQPLIIEINASPTWAPSTAVTARLCA				
GVQADTLRVVIDRLDRNCDTGAFELIYKQPVTTSPASTPGPSCLLPMYSIDTRARSD				
DSTASWNLRLPCRQARP"				
ORIGIN				
Alignment Scores:				
Pred. No.: 1..82e-77 Length: 4184				
Score: 1020.00 Matches: 220				
Percent Similarity: 55.5% Conservative: 69				
Best Local Similarity: 42.2% Mismatches: 151				
Query Match: 35.7% Indels: 81				
DB: 4 Gaps: 12				
US-10-635-977-2 (1-541) x HSM808322 (1-4184)				
Qy	54	GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrlLeuGlyGlnLeuGluHisGlu	73	
Db	2472	GAGTTTGTGGATGAAGCTCTGTGTGTCGCGAGGAGTACCTTAGCAACTTGCCCATG	2531	
Qy	74	AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu	93	
Db	2532	GACATCGACAGACCTGGAGGCCCGCTGTACTCACCCCGAGGGCTGCTCCTCTTC	2591	
Qy	94	ThrGlnGlnTyrlTyrlSerLeuValHisGlyAspAlaPheIleSerAanSerArgAsnTyrl	113	
Db	2592	CTCAGCGCTACTACCAAGTGTGTCCAGAGGGGCGAGCACTCGACACTCAG	2651	
Qy	114	PheSerGlnCysGlnAlaLeuLeuAanArgIleThrSerValAanProGlnThrAspIle	133	
Db	2652	GTCCAGCGCTGTGAGGACATCTCGACAGCTGAGCGCGGTGTACCCAGATAGATG	2711	
Qy	134	AspGlyLeuArgAanIleTrpIleIleLysProAlaAlaLysSerArgGlyArgAspIle	153	
Db	2712	GAAGGGGATCGCAACATCTGATCGTGAACCCAGGACCAAGTCCCGGACGAGGCATC	2771	
Qy	154	ValCysMetAspArgValGluGluIleLeuLeuGluLeuAlaAlaAspHisProLeuSer	173	
Db	2772	ATGTGCATGCAGCCACTGGAGGAGATGCTGAAGTGTGTGAACCGCAACCCCGTGTG	2831	


```
QY 174 ArgAspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThr 193
Db 2832 AAGGACGGCAAGTGGGTGGTGCAGAGTATATTGAGCGGCCCTCTCATCTTTGGCACC 2891
QY 194 LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
Db 2892 AAGTTTGACTCAGACAGTGGTTCCTGGTAACACTGACTGAACCACTTACCGTGTGTTTC 2951
QY 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
Db 2952 TACCGGCAGAGCTATATCCGCTTTCCACGCGAGCCCTTCTCCCTGAAGAACCTGGACAC 3011
QY 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
Db 3012 TCAGTGCACCTGTGCACCAACTCCATCCAGAGCACCTGGAGAACTCATCGCCATCGCAT 3071
QY 254 ProLeuLeuProAlaHisLeuMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArg 273
Db 3072 CCACTGCTTCCGCGCAGACACATGTGGTCTAGCCAGAGGTTCAGGCCACCTGCAGGAG 3131
QY 274 GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla 293
Db 3132 ATGGGTGCCCAATCTGGTCCACCATCATCGTCCCTGGCATGAGATGCTGTGATC 3191
QY 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
Db 3192 CACGCACCTTCCAGACCTCCAGCACACCGTGCAGTGTCCGAGGCCAGCTTTCAGCTCAT 3251
QY 314 GlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluLeuAsnSerSer 333
Db 3252 GCGCTGTACTTCTGTTCGGGGAGGACTTCCAGCCCTGGCTGATTGAGATCAACGCCAGC 3311
QY 334 ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353
Db 3312 CCCACGATGGCACCTCCACAGCATCATCTGCCGCTCTGTCTGGGTGCAAGCTGAC 3371
QY 354 ThrIleLysValAlaVal-----AspArgSerCysAspIleGlyAsnPheGlu 369
Db 3372 ACCCTGCGCGTGTCTATTACCGGATGCTGGACCGCACTGTGACACAGGAGCCTTTGAG 3431
QY 370 LeuLeuTrpArgGlnProVal---ValGluPro-ProProPheSerGlySerAspLeuCy 388
Db 3432 CTCATCTATAGCAGCCCGTCAACCACTTCCCGAGCTCCACACAGGCG----- 3480
QY 388 sValAlaGlyValSerValArgAlaArgArgGlnValLeuProValCysAsnLeuLy 408
Db 3480 ----- 3480
QY 408 sAlaSerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArgGlyProSerAlaMetPr 428
Db 3481 ----CCAGTGCTTCTCCCATGTACTCCGACACCGAGGCCAGTCTCTCAGACGACAGC 3536
QY 428 oAspProAlaGlnGlyProProSerProAlaLeuGlnArgAspLeuGlyLeuLysGluG 448
Db 3537 ACAGCAAGCTGTGTGGCACTAAGGCCCTGTGCACACACAGCAA---GGCCTTGAAGGATC 3593
QY 448 uLys-----GlyLeuProLeuAlaLeuAlaPro---LeuArgGlyAla----- 462
Db 3594 TACCCACGGCTAAGTCTTCATTCTCCCTCCCAACCGAACCTTGATTTCAAGGTGGACCCA 3653
QY 463 -----AlaGluSerGlyGlyAlaAlaGlnProThrArgThr----- 474
Db 3654 GCATCTCTGAGCCAGAGAGGTGGCCCTCGACCTGTGACTCACACCCAGTGGACAGTGTCT 3713
QY 475 -----LysAlaAlaGlyLysValGluLeuProAlaCysProCysAr 488
Db 3714 GAGCAGCGGTGAGGGCTGAGGGGACAGGAGGGGAGGTCTCCAGGCTG----- 3765
QY 488 gHisValAspSerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTr 508
Db 3766 -----CTGGACCCCAAGGAAGAGCTG 3788
```

```
QY 508 pAspProAsnGlnLeuAsn-----AlaHisProLeuGluProVal 522
Db 3789 GTCTCCCTCAGAGGCCCTTCTCCACAGACTTCTGATCATCTCTTCTCCCTCCT 3848
QY 522 uArgGlyLeuLysThrAlaGluGlyAlaLeuArgPro-----ProProGlyG 538
Db 3849 T-----TCACACCGAGGCTCTCTCTCTGTGCTCCGAGGCCCGCAGCTGG 3896
QY 538 Y 538
Db 3897 A 3897

RESULT 9
LOCUS BU937372
DEFINITION IMAGECOURT_10519180 NIH_MGC_169 Mus musculus cDNA clone
ACCESSION BU937372
VERSION BU937372.1 GI:24126191
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 775)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Dr. Jonathan Kuo, NIMH
cDNA Library Preparation: Michael Brownstein Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCW3002 row: g column: 16
High quality sequence stop: 569.

FEATURES
source
1..775
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6705256"
/lab_host="DH10B (TI-phage-resistant)"
/clone_lib="NIH_MGC_169"
/note="Organ: Testicles; Vector: pDNR-LIB; Site 1: SfiI
(ggcatatggcc); Site 2: SfiI (ggcgctcgcc); cDNA made
by oligo-dT priming and directionally cloned. 5' and 3'
adaptors were used in cloning as follows:
5'-AAGCAGTGTATCAACGAGAGTGGCATTCAGCCGGG-3' and
5'-ATTAGAGCGGAGCGGCGACATG-(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 0.5 kb
size fraction. Library created in the laboratory of M.
Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 1,25e-76 Length: 775
Score: 998.00 Matches: 189
Percent Similarity: 83.6% Conservative: 25
Best Local Similarity: 73.8% Mismatches: 38
Query Match: 35.0% Indels: 5
DB: 5 Gaps: 1

US-10-635-977-2 (1-541) x BU937372 (1-775)

QY 197 IleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGlu 216
Db 4 ATCAGACAGTGGTTCCTCTCTGTCACAGACTGGAAATCCCTTAACCATCTGCTTCTTACAAAGAA 63
```

```
QY 217 SerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHis 236
Db 64 AGGTACTCGCTTCTCCACACAGCGCTTCTCCGTGGACAAGCTGGACGCGCATCCAC 123
QY 237 LeuCysAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeu 256
Db 124 CTGTGTAAACAATCCCATCAGAGCGCTCAAAATGATAAGAGCGCATCGCTGCTTA 183
QY 257 ProAlaHisAsnMetTyrThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArg 276
Db 184 CCTTGTTCATAACATGTGGACCACTCGTTTCCAGAGTACCTGTCAGAGAAGGGGCCGA 243
QY 277 GlyAlaValTyrGlySerValIleTyrProSerMetLysLysAlaIleHisAlaMet 296
Db 244 GGAGGAACGTGGGTGAGCATCATCTACCCGCTTATGAAAAGAGCTGTCCCAATGCCATG 303
QY 297 LysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAsp 316
Db 304 AGGTAGTCCAGGACCACGTAGAACGCCGTAAAGACAGCTTCAGGCTCTACGGAGCGGAC 363
QY 317 PheValLeuGlyArgAspPheArgProTyrLeuIleGluLeuAsnSerSerProThrMet 336
Db 364 TTCATCTGNGCGGAGACTTCAGCCCATGGCTTATCGAGATCAACTCCAGCCCAACCATG 423
QY 337 HisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLys 356
Db 424 CACCCCTCCACTCCCGTACAGCTCAGCTCTGTGCCCGAGGTGCGAGGACACCATCAAG 483
QY 357 ValAlaVal-----AspArgSerCysAspIleGlyAsnPheGluLeuLeuTyr 372
Db 484 GTGGTGGTGGATCGCAAACTGGACCGCAAACTGTGACATCGGCACTTTGAGCTTCTGTGG 543
QY 373 ArgGlnProValValGluProProPheSerGlySerAspLeuCysValAlaGlyVal 392
Db 544 CGCAGCGCTCGCGTGGAGCTGCCACCGTTCAACGGGTCTGACCTCTGTGTGGAGGTATC 603
QY 393 SerValArgAlaArgArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSer 412
Db 604 AGCGTGAAGAAAGCAAGAACAAATGCCGCCATTTGCCCAATTCGCCGCTCTCAGAATCA 663
QY 413 LeuLeuAspAlaGlnProLeuLysAlaArgGlyProSerAlaMetProAspProAlaGln 432
Db 664 CTCTTAGATCTCCACCACAAAGTGGCGAGTG-CCCCGGCGCCTGTATGGAAACAGTGATC 722
QY 433 GlyProProSerProAlaLeuGlnArgAspLeuGlyLeuLysGluGlu 448
Db 723 AGGCCTCCAGAACGACCCGTGGCGGAGGACTGGGAACGGGAAGAA 770

RESULT 10
DQ045868
LOCUS
DEFINITION
1059 bp DNA linear GSS 02-JUN-2005
Homo sapiens ARPC4 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION
DQ045868
VERSION
DQ045868.1 GI:66897083
KEYWORDS
GSS.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 1059)
Nielsen,R., Bustamante,C., Clark,A.G., Glanowski,S., Sackton,T.B.,
Hubisz,M.J., Fedel-Alon,A., Tanenbaum,D.M., Civallo,D.,
White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
A Scan for Positively Selected Genes in the Genomes of Humans and
Chimpanzees
(eur) PLoS Biol. 3 (6), E170 (2005)
JOURNAL
PUBMED
15869325
REFERENCE
2 (bases 1 to 1059)
Nielsen,R., Bustamante,C., Clark,A.G., Glanowski,S., Sackton,T.B.,
Hubisz,M.J., Fedel-Alon,A., Tanenbaum,D.M., Civallo,D.,
```

White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
Direct Submission
Submitted (05-MAY-2005) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment. Translation starts at the beginning of
alignment.

FEATURES

source

1..1059

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
-1..>1059
/gene="ARPC4"
/locus_tag="HC11626"

gene

ORIGIN

Alignment Scores:

Pred. No.: 1 46e-68 Length: 1059
Score: 908.00 Matches: 165
Percent Similarity: 73.2% Conservatives: 48
Best Local Similarity: 56.7% Mismatches: 78
Query Match: 31.8% Indels: 0
DB: 11 Gaps: 0

US-10-635-977-2 (1-541) x DQ045868 (1-1059)

```
QY 74 AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu 93
Db 4 GACATCGACAAGACCTGGAGGCCCGCTGACTCACCCCGGAGGCTGGTCCCTCTTC 63
QY 94 ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
Db 64 CTCACGCGTACTACCAAGTGTGTCCAGAAAGGGCGAGAACTCAGGCACCTCGACACTCAG 123
QY 114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
Db 124 GTCCAGCGCTGTGAGACATCTTCGACAGCTGCAGCGCTGTGTACCCAGATAGACATG 183
QY 134 AspGlyLeuArgAsnIleTrpIleLysProAlaAlaLysSerArgGlyArgAspIle 153
Db 184 GAAGGGGATCGCAACATCTGGATCGTGAAGCCAGGAGCAAGTCCGCGGACGAGGCATC 243
QY 154 ValCysMetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLeuSer 173
Db 244 ATGTGATGACCACTGAGGAGGAGTGTGAAGCTGGTGAACCGCAACCCCGTGGTGATG 303
QY 174 AtgAspAsnLysTyrValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThr 193
Db 304 AAGACGGCAAGTGGGTGGTGGAGAGTATATTGAGCGGCCCTCTCTCATCTTTGGCACC 363
QY 194 LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
Db 364 AAGTTTGACCTCAGACAGTGTTCCTGGTAACTGACTGGAACCCACTTACCGTGTGGTTC 423
QY 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
Db 424 TACCGGACAGCTATATCCGCTTTTCACGACGCCCTTCTCCCTGAAGAACCTGGGACAC 483
QY 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
Db 484 TCAGTGACCTGTGCAACAACTCCATCCAGAGCAGCTCGAGAACTCATGCCATCGGCAT 543
QY 254 ProLeuLeuProAlaHisAsnMetTyrThrSerThrArgPheGlnGluTyrLeuGlnArg 273
Db 544 CCATGCTTCGCGCAGACACATGTGCTAGCCAGAGGTTCCAGGCCACCTCGCAGGAG 603
QY 274 GlnGlyArgGlyAlaValTyrGlySerValIleTyrProSerMetLysLysAlaIleAla 293
Db 604 ATGGGTGCCCAATGCTTGGTCCACCATCATCTGCTGCGCATGAAGGATGTGTGATC 663
QY 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
```

```

Db      664  CAGCGACTTCAGACCTCCAGGACACCGTCAGTGTGGAGGCCAGCTTTGAGCTCTAT 723
QY      314  GlyAlaAspPheValLeuGlyArgAspPheArgProTrrPheLeuLeuGluLeuAenSer 333
Db      724  GCGCGTACTCTGCTGTGGGGAGGACTTCAGCCCTGGCTGATTGAGATCAACGCCAGC 783
QY      334  ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353
Db      784  CCACATGGCACCCTCCACAGCAGTCACTGCCCGGCTCTGTCTGGCGTGAAGCTGAC 843
QY      354  ThrIleLysValAlaValAspArgSerCysAsp 364
Db      844  ACCCTGGCGTGTCAATTGACCGATGCTGGAC 876

```

RESULT 11

```

BM808637 1023 bp mRNA linear EST 05-MAR-2002
LOCUS AGENCOURT_6617786 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734232
DEFINITION 5', mRNA sequence.

```

```

ACCESSION BM808637.1 GI:19125460
VERSION EST.
KEYWORDS
SOURCE Homo sapiens (human)

```

ORGANISM

```

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

```

```

1 (bases 1 to 1023)

```

REFERENCE

```

AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

```

COMMENT

```

Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Invitrogen
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

```

```

Plate: LLAM12739 row: d column: 09

```

```

High quality sequence start: 6

```

```

High quality sequence stop: 664.

```

FEATURES

source

```

1..1023
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5734232"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH MGC 124"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site 1: EcoRV
(destroyed); Site 2: NotI; RNA source male hippocampus,
age 27. Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.4 kb, insert size range 0.9-4 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 012."

```

ORIGIN

```

Alignment Scores:
Pred. No.: 4 238-67 Length: 1023
Score: 891.00 Matches: 177
Percent Similarity: 67.7% Conservative: 54
Best Local Similarity: 51.9% Mismatches: 94
Query Match: 31.2% Indels: 16
DB: 3 Gaps: 4

```

```

US-10-635-977-2 (1-541) x BM808637 (1-1023)

```

```

BM808516
LOCUS
DEFINITION

```

```

BM808516 1022 bp mRNA linear EST 05-MAR-2002
AGENCOURT_6617783 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734160
5', mRNA sequence.

```

```

QY      107  IleSerAsnSerArgAsnTyrPheSerGln-----CysGlnAlaLeuLeuAenArg 123
Db      14   GTCCCGCGCAGCAGGACCGCTCTACACTCAGGTCCAGCGCTGTGAGGACATCTCTCGACGAG 73
QY      124  IleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrrIleLys 143
Db      74   CTGACGCGCTGGTACCCAGATAGACATGGAAGGGGATCGCAACATCTGGATCGTGAAG 133
QY      144  ProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGluLeu 163
Db      134  CCAGGAGCCCAAGTCCCGCGGACGAGCATCATGTGATGACCACTGAGGAGATGCTG 193
QY      164  GluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrrValValGlnLysTyr 183
Db      194  AAGCTGCTGAACGGCAACCCCGTGTGTATGAAGGACGGCAAGTGGTGTGTCAGAAAGTAT 253
QY      184  IleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrrPheLeuVal 203
Db      254  ATTGAGCGGCCCTCTCTCATCTTTGGCACCAAGTTGACCTCAGACAGTGGTTCCTGGTA 313
QY      204  ThrAspTrrAsnProLeuThrIleTrrPheTyrLysGluSerTyrLeuArgPheSerThr 223
Db      314  ACTGACTGGAAACCCACTTACCGTGTGGTCTTACCGGACAGCTATATCCGCTTTCCACG 373
QY      224  GlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAlaValGln 243
Db      374  CAGCCCTTCTCCCTGAAGAACCTGGCAACTCAGTGCACCTGTGCAACAACTCCATCCAG 433
QY      244  LysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMetTrrThr 263
Db      434  AAGCACCTCGGAACTCATGCCATCGGCATCCACTGCTTCCGCCAGACAACTGTGGTCT 493
QY      264  SerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTrrGlySerVal 283
Db      494  AGCCAGAGGTTCCAGGCCACCTCGCAGGAGATGGGTGGCCCAAAATGCTTGGTCCCATC 553
QY      284  IleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAspHisVal 303
Db      554  ATCGTGCTGGCATCAAGGATGCTGTGATCCACGCACCTTCAGACCTCCCGCAGCACCGTG 613
QY      304  GluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArgAspPhe 323
Db      614  CAGTGTGGAAAGGCCAGCTTTGAGCTCTATGCGCTGACTTCTGCTGGGAGGACTTC 673
QY      324  ArgProTrrPheLeuLeuGluLeuLeuLeuGlnArgGlnGlyArgGlyAlaValTrrGlySerVal 343
Db      674  CAGCCCTGGCTGATTGAGATCAACGCCAGCCACCATGGCCCTCCACACGACGATCACT 733
QY      344  AlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAspArg----- 361
Db      734  GCCCGGCTCTGTGCTGGCGTGCAGAGCTGACACCCCTGCGTGTCTATTGACCGGATGCTG 793
QY      362  -----SerCysAspIleGly-AsnPheGluLeuLeuTrrPargGlnProVal---ValG 378
Db      794  GACCGCGACTGTGACCCAGGAGCCCGTTGAGCTCATCTATAGCAGCCCGTAACCATCTTC 853
QY      378  upro-ProPheSerGlySerLeuCysValAlaGlyValSerValArgAlaAa 398
Db      854  CCCAGCTTCCCAAGGCCCGCCAGCTTGTCTTCTCCCATGTATCTCCGAACACACAGGGG 913
QY      398  rGArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnP 418
Db      914  CCAGG-----TCCCTCAGAACGACAGCCACAGCAAGGCTGGTGGGGG 955
QY      418  roLeuLysAlaArgGlyProSerAlaMetProAspPrrProAlaGlnGlyProProSerPro 437
Db      956  AACTAAGTGCCTTGGTCGACCCACTGGGCAAGGGCCCTTGTAGGGGAGCTCCTTACCCA 1014

```

```

RESULT 12

```

```

BM808516

```

```

LOCUS
DEFINITION

```

```

BM808516 1022 bp mRNA linear EST 05-MAR-2002
AGENCOURT_6617783 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734160
5', mRNA sequence.

```

```

ACCESSION   BM808516
VERSION     BM808516.1  GI:19125339
SOURCE      EST.
ORGANISM    Homo sapiens (human)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.
REFERENCE   1 (bases 1 to 1022)
            NIH-MGC http://mgs.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-f@mail.nih.gov
            Tissue Procurement: Invitrogen
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LAM12739 row: a column: 09
            High quality sequence start: 78
            High quality sequence stop: 658.
            Location/Qualifiers
FEATURES             1..1022
     source           /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5734160"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH MGC 124"
                     /notes="Organ: brain; Vector: pCMV-SPORT6; Site 1: EcoRV
                     (destroyed); Site 2: NotI; RNA source male hippocampus,
                     age 27. Library is oligo-dT primed and directionally
                     cloned (EcoRV site is destroyed upon cloning). Average
                     insert size 1.4 kb, insert size range 0.9-4 kb. Library is
                     normalized and enriched for full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 012."
ORIGIN
Alignment Scores:
Pred. No.:      4.13e-60      Length:      1022
Score:          811.00        Matches:    162
Percent Similarity: 64.2%      Conservative: 48
Best Local Similarity: 49.5%    Mismatches: 72
Query Match:    28.4%          Indels:     46
DB:             3             Gaps:       5

US-10-635-977-2 (1-541) x BM808516 (1-1022)

Qy 116 GlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGly 135
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 41 CGCTGTGAGGACATCTCTCGACAGCAGCTGCGCGCGTGTATCCCAAGATAGACATGGAAGGG 100
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy 136 LeuArgAsnIleTrrIleLysProAlaLysSerArgGlyArgAspIleValCys 155
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 101 GATCCCAACATCTGATCGTGAACAGGACAGGACCAAGTCCCGGACGAGGCATCATGTGC 160
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy 156 MetAspArgValGluGluLeuLeuAlaAlaAspHisProLeuSerArgAsp 175
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 161 ATGGACCACTGGAGGAGATCTGAAGCTGTGTGAACCAACCCCGTGTGTGATGAAGGAC 220
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy 176 AsnLysTrpValGlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPhe 195
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 221 GGCAAGTGGTGGTGGCAGAGATATATTGAGCGGCCCTCTCTCATCTTTGGCACCAGTTT 280
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy 196 AspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLys 215
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 281 GACCTCTTACAGTGGTCTCTGTTAACTGACTGGAAACCCACTTACCGTGTGTCTTACCGC 340
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

```

Qy 216 GluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIle 235
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 341 GACAGCTATATCGCTTTTCCAGGAGCCCTTCTCCCTGAAGAACTGGACAACTCAGTG 400
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy 236 HisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeu 255
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 401 CACCTGTGCAACAACACTCCATCCAGAAACCTGGAGAACTCATGCCATCGGCATCCACTG 460
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy 256 LeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGly 275
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 461 CTTCCGCCACAGACAACATGTGTGTAGCCAGAGGTTCCAGGCCACCTGCGAGGAGATGGGT 520
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy 276 ArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysLysAlaHisAla 295
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 521 GCCCCAAATGCTTGGTCCACCATCATCTGTCCTTGGCATGAAGATCGTGTGATCCAGCA 580
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy 296 MetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAla 315
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 581 CTTTCAGACCTCCAGGACACCGTGCAGTGTGCGAAGCCAGCTTTGAGCTCTATGGCGCT 640
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy 316 AspPheValLeuGlyArgAspPheArgProTrpLeuIleGluLeuAsnSerSerProThr 335
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 641 GACTTCGTGTTCGGGAGGACTTCCAGCCCTGCTGATTGAGATCAACGCCGCCCCACG 700
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy 336 MethisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIle 355
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 701 ATGGCAACCCCTCACAGCATCTGCGCGGCTCTGTGGCGTGCAGCTGACACCCCTG 760
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy 356 -LysValAlaValAspArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnPr 375
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 761 CCGCGTGGCATTTGGCGGGAATGCC-----TGGACCCGCAA 796
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy 375 oValValGluProProPheSerGlySerAspLeuCysValAlaGlyValSerValAr 395
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 797 TGTGACCC--CCGGAACCCCTTT----- 816
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy 395 gArgAlaArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAs 415
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 817 -----GA 818
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy 415 pAlaGlnProLeuLys---AlaArgGlyProSerAlaMet---ProAspProAlaGlnGl 433
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 819 ACCTCATCTCTAATAAGTAAGCCGCGGACCACTTCTCTCAGCGCTCCCGCCACCAAGGG 878
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy 433 YPro-----ProSerPro 437
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 879 CCCAGCTTGGCTTCCGCC 897
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

RESULT 13
DQ045869
LOCUS
DEFINITION
DQ045869
Pan troglodytes ARPC4 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
DQ045869.1 GI:66897084
GSS.
Pan troglodytes (chimpanzee)
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Pan
1 (bases 1 to 1059)
Nielsen,R., Bustamante,C., Clark,A.G., Glanowski,S., Sackton,T.B.,
Hubisz,M.J., Fledel-Alon,A., Tanenbaum,D.M., Civallo,D.,
White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
A scan for Positively Selected Genes in the Genomes of Humans and
Chimpanzees
(er) PLOS Biol. 3 (6), E170 (2005)
JOURNAL
PUBMED
15869325
REFERENCE
2 (bases 1 to 1059)
Nielsen,R., Bustamante,C., Clark,A.G., Glanowski,S., Sackton,T.B.,
Hubisz,M.J., Fledel-Alon,A., Tanenbaum,D.M., Civallo,D.,
White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.

```

TITLE Direct Submission
JOURNAL Submitted (05-MAY-2005) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment. Translation starts at the beginning of alignment.

FEATURES

source
1..1059
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1059
/gene="ARPC4"
/locus_tag="HC11626"

ORIGIN

Alignment Scores:
Pred. No.: 4,35e-60 Length: 1059
Score: 811.00 Matches: 154
Percent Similarity: 67.0% Conservative: 41
Best Local Similarity: 52.9% Mismatches: 96
Query Match: 28.4% Indels: 0
DB: 11 Gaps: 0

US-10-635-977-2 (1-541) x DQ045869 (1-1059)

Qy 74 AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu 93
Db 4 GACATCGACCAAGGACCTGGAGGCCACCTGACCTACCTACCCCAAGGCTGGTCCCTCTTC 63
Qy 94 ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
Db 64 CTCACGCGCTACTACCAAGTGTCTCCAGGAAGGGGAGAACTCAGGCACCTCCGACACTCAG 123
Qy 114 PheSerGlnCysGlnAlaLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
Db 124 GTCCAGCGCTGTGAGGACATCTCGACGACGTGGCGGCGGTGTACCCAGATAGCATG 183
Qy 134 AspGlyLeuArgAsnIleTrpIleIleLysProAlaAlaLysSerArgGlyArgAspIle 153
Db 184 GAAGGGATCGCAACACTGTGATCTGGAAGCAGGAGCAAGTCCCGGGGAGGAGCATC 243
Qy 154 ValCysMetAspArgValGluIleLeuGluLeuAlaAlaAspHisProLeuSer 173
Db 244 ATGTGCATGGACCACTCGAGGAGATGTGAAGCTGTGAAGCAAGCCCGCTGGTGATG 303
Qy 174 ArgAspAsnLysTrpValGlnLysTyrIleGluThrProLeuLeuIleCysAspThr 193
Db 304 AAGGACGGCAAGTGGTGGTGAGAGATATATTGAGCGGCGCCCTCTCATCTTTGGCAC 363
Qy 194 LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
Db 364 AAGTTTGACCTCAGACAGTGTCTCTGTAACTGACTGGACCCCACTTACCGTGTGGTTC 423
Qy 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
Db 424 TACCGCGACAGTATATCCGCTTTTCCAGCGAGCCCTCTCTCCCTGAAGAACTGGCAAN 483
Qy 234 AlalIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
Db 484 NNN 543
Qy 254 ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArg 273
Db 544 CCACGTGCTTCGCCAGACAAACATGNNNTCTAGCCGAGGTTCAGGCCACCTGCAGGAG 603
Qy 274 GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla 293
Db 604 ATGGGTGCCCAAAATGCTTGTGTCCACCATCATCTCGTCCGCTGGCATGAAGCATGCTGTGATC 663
Qy 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
Db 664 CACGACCTTACAGCTCCCAAGACACCGCTGCAGTGTGGAAGGCCAGCTTTTGGCTCTAT 723

Qy 314 GlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluIleAsnSerSer 333
Db 724 GCGCGTGACTTCGTGTGGGGAGGACTTCCAGCCCTGGCTGATTGAGATCAACGCCAGC 783
Qy 334 ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353
Db 784 CCCAGATGGCACCTCCACAGCAGTCACTGCCGCTCTGTGCTGGCGTGAAGCTGAC 843
Qy 354 ThrIleLysValAlaValAspArgSerCysAsp 364
Db 844 ACCCTGCGCTGCTCATTGACCGGATGCTGGAC 876

RESULT 14

CK597795 777 bp mRNA linear EST 22-JAN-2004
LOCUS AGENCOURT 17837841 NIH_MGC_238 Rattus norvegicus cDNA clone
DEFINITION IMAGE:7135519 5', mRNA sequence.

CK597795

CK597795.1 GI:41110910

EST.

Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Rattus.

NIH-MGC http://mgi.nci.nih.gov/

1 (bases 1 to 777)

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Daniela S. Gerhardt, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Blgd. 31 Rm10A07 Bethesda, MD 20892

Email: cgapbs-remail.nih.gov

Tissue Procurement: Howard Jacobs

cDNA Library Preparation: Express Genomics

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM15042 row: 0 column: 05

High quality sequence stop: 693.

Location/Qualifiers

1..777

/organism="Rattus norvegicus"

/mol_type="mRNA"

/db_xref="taxon:10116"

/clones="IMAGE:7135519"

/tissue_type="testis, pooled"

/lab_host="DH10B TON"

/clone_lib="NIH_MGC_238"

/note="Organ: testis; Vector: pExpress-1; Site: 1: EcoRV;

Site 2: NotI; RNA obtained from testis tissue of 8 wk old

animal. Tissues were snap-frozen and kept at -80C before

RNA extraction and purification (tri-reagent method). cDNA

was primed using oligo-dT primer:

5'-pGATAGTTCATGATCGGAGCGGCCCTTC-3' and cloned into

the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb

resulted in an average insert size of 1.9 kb. This primary

library is normalized (non-normalized primary library is

NIH MGC 237) and was constructed by Express Genomics

(Frederick, MD)"

Alignment Scores:

Pred. No.: 1.14e-58 Length: 777

Score: 792.50 Matches: 156

Percent Similarity: 82.0% Conservative: 31

Best Local Similarity: 68.4% Mismatches: 36

Query Match: 27.8% Indels: 6

DB: 7 Gaps: 3

ORIGIN

Alignment Scores:
Pred. No.: 1.14e-58 Length: 777
Score: 792.50 Matches: 156
Percent Similarity: 82.0% Conservative: 31
Best Local Similarity: 68.4% Mismatches: 36
Query Match: 27.8% Indels: 6
DB: 7 Gaps: 3

Db 543 ACCATGGCGCCCTCCACGGCGGTCACTGCCCCGGCTCTGTGCTGGAGTGCAGCTGACACA 602
Qy 355 IleLysValAlaVal-----AspArgSerCysAspIleGlyAsnPheGluLeu 370
Db 603 CTGGTGTGTGTCATCGACCGCGGACTGGACCGCAACTGTGTGACACGGGAGCCTTTGAGCTC 662
Qy 371 LeuTyrArgGlnPro 375
Db 663 ATCTACAAGCAGCCT 677

Search completed: April 4, 2006, 11:34:23
Job time : 6008 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2006, 09:36:06 ; Search time 779 Seconds
(without alignments)
4628.496 Million cell updates/sec

Title: US-10-635-977-2

Perfect score: 2854

Sequence: 1 MASSILKXWVSHQSCSRSSR.....LRGLTKAEGALRPPPGKGKS 541

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/abs/ABSSWEB spool/US10635977/runat_04042006_103529_9972/app query.fasta.1
-DB=N Geneseq -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss03p
-USER=US10635977 @CGN_1_1.1096 @runat_04042006_103529_9972 -NCPU=6 -ICPU=3
-NO WMAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_21.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*
14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2854	100.0	1838	12	ADJ93357 Human BGS
2	2843	99.6	3554	12	ADJ93364 Human tub
3	2698.5	94.6	1859	12	ADJ93362 Human BGS
4	2693	94.4	3465	12	ADJ93363 Human BGS

c	5	2689	94.2	1939	12	ADJ93361	Adj93361 Human BGS
	6	1802	63.1	101270	12	ADQ17814	Adq17814 Human sof
	7	1299	45.5	726	12	ADJ93359	Adj93359 Human BGS
	8	1070.5	37.5	2380	6	AAS99894	Aas99894 Polynucle
	9	1054	36.9	2326	4	AAH16735	Aah16735 Human cdn
	10	1020	35.7	2538	7	ADS73168	Ads73168 Human kid
	11	1020	35.7	2538	7	ADM42022	Adm42022 cDNA elev
	12	1020	35.7	2553	13	ACN37881	Acn37881 Tumour-as
	13	1020	35.7	2881	3	AAC77214	Aac77214 Human ORF
	14	1020	35.7	3001	4	AAI58606	Aai58606 Human pol
	15	1020	35.7	3001	5	ADO98824	Ado98824 DNA encod
	16	1020	35.7	3001	5	ADB48584	Adb48584 Novel hum
	17	1012	35.5	3294	8	ABX34502	Abx34502 Human mdd
	18	998	35.0	2611	3	AAF21812	Aaf21812 Human bre
	19	983.5	34.5	2848	11	ADM03081	Adm03081 Human cdn
	20	940.5	33.0	5282	12	ADQ64426	Adq64426 Novel hum
	21	911.5	31.9	2412	12	ADH45468	Adh45468 Human mol
	22	897	31.4	2469	12	ADH45478	Adh45478 Human mol
	23	781.5	27.4	2217	10	ACA92443	Ac92443 DNA encod
c	24	772	27.0	3828	13	ADR07582	Adr07582 Full leng
	25	760	26.6	490	12	ACH91699	Ach91699 Human gen
	26	733	25.7	2979	4	ABL17657	Ab117657 Drosophil
	27	556.5	19.5	5728	4	ABL17656	Ab117656 Drosophil
	28	536.5	18.8	2543	4	ABL17449	Ab117449 Drosophil
	29	519.5	18.2	2044	4	AAI60392	Aai60392 Human pol
	30	516	18.1	1206	13	ADU02015	Adu02015 Novel hum
	31	514.5	18.0	4615	4	ABL17448	Ab117448 Drosophil
	32	489.5	17.2	1567	8	ACC46592	Acc46592 Human dit
	33	480.5	16.8	1958	4	AAH18689	Aah18689 Human cdn
c	34	478.5	16.8	10468	4	AAK77961	Aak77961 Human imm
	35	405	14.2	418	8	ABX52103	Abx52103 Bovine ES
	36	394	13.8	1728	5	AAS70158	Aas70158 DNA encod
	37	390.5	13.7	2250	5	AAS69383	Aas69383 DNA encod
	38	381.5	13.4	1085	4	AAF63819	Aaf63819 Human sec
	39	336	11.8	1579	8	ABX05228	Abx05228 Human nov
	40	328	11.5	1571	4	ABL14515	Ab114515 Drosophil
	41	320	11.2	2259	12	ADQ64293	Adq64293 Novel hum
	42	318	11.1	2417	10	ADC30440	Adc30440 Human nov
	43	307.5	10.8	3141	4	ABL06959	Ab106959 Drosophil
	44	305	10.7	5314	4	AAI58261	Aai58261 Human pol
	45	305	10.7	5314	5	ADQ98468	Adq98468 DNA encod

ALIGNMENTS

RESULT 1

ADJ93357

ID ADJ93357 standard; cDNA; 1838 BP.

XX

AC ADJ93357;

XX

DT 06-MAY-2004 (first entry)

XX

DE Human BGS-42 cDNA sequence SeqID1.

XX

testis-specific tubulin tyrosine-ligase-like polypeptide;
BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
osteopathic; nootropic; antiparkinsonian; antiarthritic; antitachytic;
anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
brain cancer; liver cancer; proliferative condition; testis; lung;
small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
sepsis; acne; Sjogren's disease; scleroderma; human; gene; ss.

XX Homo sapiens.

OS

XX

FH

Location/Qualifiers

FT CDS 153..1778
FT /*tag= a
FT /product= "Human BGS-42 protein"
XX
XX WO2004005487-A2.
XX
XX 15-JAN-2004.
XX
XX 09-JUL-2003; 2003WO-US021605.
XX
XX 09-JUL-2002; 2002US-0394725P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Feder JN, Wu S, Nelson TC;
XX
XX WPI; 2004-099381/10.
XX P-PSDB; ADJ93358.
XX

PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.

XX Claim 1; SEQ ID NO 1; 343pp; English.

XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, neurotropic,
CC antiparkinsonian, antiarthritic, antiaesthetic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a cDNA
CC which encodes the human BGS-42 protein of the invention.

XX
XX Sequence 1838 BP; 381 A; 576 C; 583 G; 298 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.33e-176 Length: 1838
Score: 2854.00 Matches: 541
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 12 Gaps: 0

US-10-635-977-2 (1-541) x ADJ93357 (1-1838)

Qy 1 MetAlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerArg 20
Dy 153 ATGGCATCCAGCATCCCTCAATGGGTGGTCCAGCCAGAGCTGCAGCAGGAGCAGAGA 212
Qy 21 SerLysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerArgGlnAsp 40
Dy 213 AGCAAGCCAGGAGCAGAGGGAGGAGCCGGGAGCAGCGACCTGACAGCAGGCAAGAT 272

Qy 41 AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys 60
Dy 273 GCTGAAATGCTGAGGCAAGCTCAGGGCCCTCCGGGGCAGCTTGTGACATCGGTGC 332
Qy 61 LysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp 80
Dy 333 AAGGTGTGCCAGSCCTACCTGGGCGAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT 392
Qy 81 AlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrTrpSerLeu 100
Dy 393 GCCGTGGAGACCTCACTCAGGCCGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC 452
Qy 101 ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu 120
Dy 453 GTTCATGGCGATGCTTTCATCTCCAAATTCAGAAATTAATTTTCGAGTGCAGGCTCTG 512
Qy 121 LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrp 140
Dy 513 CTGAATAGAATCAGCTCTGTGAACCTCAGACGGACATTCAGCGGCTCCGGAACATCTGG 572
Qy 141 IleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
Dy 573 ATTATAAGCCCGCGCCCAAGTCCGGGGCCGAGACATAGTGTGCATGACCGTGTGGAG 632
Qy 161 GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValVal 180
Dy 633 GAGATCCTGGAGCTGGCAGCTGCAGACCACTCTTTCCAGGGACAAAGTGGGTGTC 692
Qy 181 GlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrp 200
Dy 693 CAGAGTACATCAGACGCGCTGCTCATCTGTGACCAACCAAGTTCGACATCAGACAGTGG 752
Qy 201 PheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArg 220
Dy 753 TTCTCTCGTCACGACTGGAAACCCCTGACCATCTGGTTCTTACAGAGAGATTAATTTGCGG 812
Qy 221 PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn 240
Dy 813 TTCTCAACTCAGCGCTTCTCCCTGGCAAGCTGGACAGCGCCATCCACCTGTGCAACAAC 872
Qy 241 AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn 260
Dy 873 GCGCTCCAGAGTACCTGAAGAATGATGTGGCGCGCAGCCCTGCTGCGCGCACACAAC 932
Qy 261 MetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTrp 280
Dy 933 ATGTGACACGACCAACAGGTTCCAGGAGTACCTGACGCGCAGGCGCGTGGCGCGTGG 992
Qy 281 GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln 300
Dy 993 GGCAGCGTCAATCTACCCGCTCCATGAAGAGGCCATCGCCACGCCATGAAGTGGCCAG 1052
Qy 301 AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly 320
Dy 1053 GACCACGTGGAGCTCGAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGCTCTGGG 1112
Qy 321 ArgAspPheArgProTrpLeuIleGluLeuAsnSerSerProThrMetHisProSerThr 340
Dy 1113 AGGGACTTCAGGCGCTTGGCTGATCGAGATCAATTTCCAGCCCCCACCACCGCTCCAGC 1172
Qy 341 ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp 360
Dy 1173 CCGGTTCAGGCCCGAGCTGTGTGCAGAGTGCAGAGGACACCATCAAGTGGCGGTGGAC 1232
Qy 361 ArgSerCysAspIleGlyAsnPheGluLeuTrpArgGlnProValValGluProPro 380
Dy 1233 CGCAGCTGTGACATCGCAACTTCGAGCTCTGTGGAGGCGCGGTGGTTGAGCCGCC 1292
Qy 381 ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaArgArgGln 400
Dy 1293 CCATTACGGGGTCCGACCTCTCGTGGCGGGCGTCAAGTGTGAGGAGGAGCAGAGGAG 1352
Qy 401 ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys 420

Db 1353 GTGCTCCGCTGTCACCTCAAGGCTCGCCCTCGCTGTGGACCGCGCTGAAG 1412
 QY 421 AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln 440
 Db 1413 GCACGGGGCCCTCGGCCATGCGACACCTGCGCCAGGGACCCCATCACAGCTCTCCAG 1472
 QY 441 ArgAspLeuGlyLeuGlyGluGlyGlyLeuProLeuAlaLeuAlaProLeuArg 460
 Db 1473 CGGGACTTGGGACTGAAGAAGAGAGGGGCTCCCTCGCTGTGGCAGCCCTTAAGG 1532
 QY 461 GlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal 480
 Db 1533 GGGGACGCGAGCGGTGGAGCGGCACAGCCGCCACCAAGCTCTGCGAAGGTG 1592
 QY 481 GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro 500
 Db 1593 GAGTCCCGGCTCGCCCTGTGCGCACGTGGACAGTCAGGCCCAACACCGGTGTCGCC 1652
 QY 501 ValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGluPro 520
 Db 1653 GTAGCCACCGCCGCAAAAGCTGGATCCAAACCAAGCTAAATGCGCACCGCTGGAGCT 1712
 QY 521 ValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGlyLysGly 540
 Db 1713 GTGTCGGGGCTGAAGACAGAGGGCGCGCTCGTCCGCGCCCGGAGAAAGT 1772
 QY 541 Ser 541
 Db 1773 TCA 1775

RESULT 2

ADJ93364

ID ADJ93364 standard; DNA; 3554 BP.

XX AC ADJ93364;

XX DT 06-MAY-2004 (first entry)

XX DE Human tubulin tyrosine ligase protein consensus gene sequence SeqID12.

KW testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytostatic; respiratory-Gen; gastrointestinal-Gen;
 KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
 KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human; gene; ds.
 XX Homo sapiens.
 OS Synthetic.

FH Key Location/Qualifiers
 FT CDS 1695..3320
 FT /*tag= a
 FT /product= "Human tubulin tyrosine ligase protein"

PN WO2004005487-A2.

XX 15-JAN-2004.

XX 09-JUL-2003; 2003WO-US021605.

XX 09-JUL-2002; 2002US-039472P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX

PI Feder JN, Wu S, Nelson TC;

XX WPI; 2004-099381/10.

DR P-PSDB; ADJ93365.

XX

PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
 useful for preventing, treating or ameliorating a medical condition, e.g.
 PT aberrant cellular proliferation, reproductive disorders or testicular
 disorders.

XX

PS Example 4; SEQ ID NO 12; 343pp; English.

XX

CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
 -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytostatic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
 CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 CC immunosuppressive, antiseborrheic or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of the
 CC tubulin tyrosine ligase protein consensus gene sequence which was used in
 CC the exemplification of the invention.

SQ Sequence 3554 BP; 692 A; 1090 C; 1157 G; 615 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.49e-175 Length: 3554
 Score: 2843.00 Matches: 539
 Percent Similarity: 99.6% Conservatives: 0
 Best Local Similarity: 99.6% Mismatches: 2
 Query Match: 99.6% Indels: 0
 DB: 12 Gaps: 0

US-10-635-977-2 (1-541) x ADJ93364 (1-3554)

QY 1 MetAlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerArg 20
 Db 1695 ATGGCATCCAGCATCTCTCAAGTGGTGTGTGTCAGCCACAGAGCTGCAGGAGCAGCAGA 1754

QY 21 SerLysProArgAspGlnArgGluAlaGlySerSerAspLeuSerSerArgGlnAsp 40
 Db 1755 AGCAAGCCAGGGACAGAGGAGAGGCCCGGAGCAGCTGAGCAGGAGGCAAGAT 1814

QY 41 AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys 60
 Db 1815 GCTGAATGCTGAGGCAAAAGCTCAGGGGCGCTCCCGGGGAGCTTGTGACATCGGTGC 1874

QY 61 LysValCysGlnAlaTyLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp 80
 Db 1875 AAGGTGCCAGGCGCTACCTGGGCGAGCTGGAGCATGAGGACATGCACACGTCAGCAGAT 1934

QY 81 AlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlyTyrSerLeu 100
 Db 1935 GCCGTGGAGGACCTCAGTGAGGCGGAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTC 1994

Qy	101	ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu	120
Db	1995	GTTCATGGCGATGCTTTTCATCTCCATTCAAGAAATTTACTTTTCGAGTCGCCAGGCTCTG	2054
Qy	121	LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrp	140
Db	2055	CTGAATAGAAATCAGCTCTGTGAACCCCTCAGACGGACATTACCGGGCTCCGGAAACATCTGG	2114
Qy	141	IleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu	160
Db	2115	ATTATAAAGCCCGCGGCAAGTCCCGGGGCGAGACATAGTGTGCATGGACCGGTGTGGAG	2174
Qy	161	GluIleLeuGluLeuAlaAlaAlaAspHisProLeuSerArgAspAsnLysTrpValVal	180
Db	2175	GAGATCTGGAGCTGGCAGCTGCAGACCAACCCCTCTTTCCAGGACACAAGTGGGTGGTC	2234
Qy	181	GlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrp	200
Db	2235	CAGAAGTACATCGAGACGCGCTCTCATCTGTGCACCAAGTTCGACATCAGACAGTGG	2294
Qy	201	PheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArg	220
Db	2295	TTCTCTGTCAGGACTGGAAACCCCTGACCATCTGGTTCTACAGGAGAGTTACTTCCGG	2354
Qy	221	PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn	240
Db	2355	TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACAAC	2414
Qy	241	AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn	260
Db	2415	GCCGTCCAGAAGTACCTGAAGATGATGTGGCGCGACGCCCTCTGTCGCCGACACACAC	2474
Qy	261	MetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTrp	280
Db	2475	ATGTGGACACGACACAGGCTTCAGAGTACCTGCAGCGCCAGGGCCCGTGGCGCTGGG	2534
Qy	281	GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln	300
Db	2535	GGCAGCGTCATCTACCCGTCATGAAGAGGCCATCGCCACGCCCATGAAGTGGCCCGAC	2594
Qy	301	AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly	320
Db	2595	GACCACGTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGGTGACTTCGTCTTGGG	2654
Qy	321	ArgAspPheArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThr	340
Db	2655	AGGGACTTCAGGCCCTGGCTGATCGAGATCAATTTCCAGGCCCCACATGCACCGCTCCAG	2714
Qy	341	ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp	360
Db	2715	CCGCTCAGGCGCCAGCTGTGTGCACAGGTGCAGAGGACACCATCAAGTGGCGGTGGAC	2774
Qy	361	ArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValValGluProPro	380
Db	2775	CGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGTGGAGGACGCGCGTGTGAGCCGCC	2834
Qy	381	ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgArgAlaArgArgGln	400
Db	2835	CCATTACGGGTCGACCTCTGGGTGGCGGGCGTCAGTGTGAGAGAGCCAGAGGGCGAG	2894
Qy	401	ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys	420
Db	2895	GTGCTGCCCGTCTGCAACTTCAGGCCTCGGCCCTCGCTGTGGACGCGCACCGCTGAAG	2954
Qy	421	AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln	440
Db	2955	GCACGGGGCCCTCTGGCCATGCCAGACCCTCGCCAGGACCCCCATCCACAGCTTCCAG	3014
Qy	441	ArgAspLeuGlyLeuLysGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuArg	460
Db	3015	CGGGACTTGGGACTGAAGAGAGAGAGGGGCTCCCCCTGGGCTTCTCTGGCACCCCTTAGG	3074
Qy	461	GlyAlaAlaGluSerGlyGlyValAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal	480

Db	3075	GGGGCAGCCGAGAGCGGTGGAGCGGCACAGCCACCCGACCAAGCTGCTGGAGGTG	3134
Qy	481	GluleuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro	500
Db	3135	GAGTCTCCGGCGTCCCTGTGCGCACGTGGACAGTCAAGCCCAACACCGGTGTCCCC	3194
Qy	501	ValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGluPro	520
Db	3195	GTAGCCAGCCCGCCAAAGCTGGGATCCAAACCCAGCTAAATGAGCACCCGCTGGAGCCT	3254
Qy	521	ValLeuArgGlyLeuLysThrAlaGluClyValAlaLeuArgProProGlyGlyLysGly	540
Db	3255	GTGCTGCGGAGCCTGAAGACAGCAGAGGGCGCGTGCCTCGCCGCCCGAGGAGAAAGGT	3314
Qy	541	Ser	541
Db	3315	TCA	3317
RESULT 3			
ID	ADJ93362		
XX	ADJ93362 standard; DNA; 1859 BP.		
AC	ADJ93362;		
XX			
DT	06-MAY-2004 (first entry)		
XX			
DE	Human BGS-42 protein-related DNA clone B SeqID10.		
XX			
XX	testis-specific tubulin tyrosine-ligase-like polypeptide;		
KW	BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;		
KW	neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;		
KW	osteopathic; nootropic; antiparkinsonian; antiarthritic; antiaesthmic;		
KW	anti-HIV; antibacterial; immunosuppressive; antiseborrheic;		
KW	dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;		
KW	tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;		
KW	testicular disorder; testicular cancer; pulmonary disorder; lung cancer;		
KW	gastrintestinal disorder; colon cancer; stomach cancer; neural disorder;		
KW	brain cancer; liver cancer; proliferative condition; testis; lung;		
KW	small intestine; brain; lymph tissue; infertility; Cushing's syndrome;		
KW	emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;		
KW	Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;		
KW	sepsis; acne; Sjogren's disease; scleroderma; human; ds.		
OS	Homo sapiens.		
XX			
PN	WO2004005487-A2.		
XX			
PD	15-JAN-2004.		
XX			
PF	09-JUL-2003; 2003WO-US021605.		
XX			
PR	09-JUL-2002; 2002US-0394725P.		
XX			
PA	(BRIM) BRISTOL-MYERS SQUIBB CO.		
XX			
PI	Feder JN, Wu S, Nelson TC;		
XX			
DR	WPI; 2004-099381/10.		
XX			
PT	New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,		
PT	useful for preventing, treating or ameliorating a medical condition, e.g.		
PT	aberrant cellular proliferation, reproductive disorders or testicular		
PT	disorders.		
XX			
PS	Example 4; SEQ ID NO 10; 343pp; English.		
XX			
CC	This invention relates to a novel testis-specific tubulin tyrosine-ligase		
CC	-like polypeptide, designated the BGS-42 polypeptide. The invention may		
CC	be useful for the development of compounds with a cytostatic, respiratory		
CC	-Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,		
CC	antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,		
CC	antiparkinsonian, antiarthritic, antiaesthmic, anti-HIV, antibacterial,		

CC immunosuppressive, antiseborrheic or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of a DNA
 CC clone sequence which is related to the invention.

XX
 SQ Sequence 1859 BP; 418 A; 563 C; 562 G; 316 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,81e-166 Length: 1859
 Score: 2698.50 Matches: 522
 Percent Similarity: 96.5% Conservative: 0
 Best Local Similarity: 96.5% Mismatches: 4
 Query Match: 94.6% Indels: 17
 DB: 12 Gaps: 2

US-10-635-977-2 (1-541) x ADJ93362 (1-1859)

CC	1	MetAlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerArg	20
QY			
DB	47	ATGGCATCAGCATCTCAAGTGGTGGTTCAGCCACCAGAGCTGCAGCAGGAGCAGCAGA	106
QY			
DB	21	SerLysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerArgGlnAsp	40
QY			
DB	107	AGCAAGCCAGGAGCCAGAGGGAGGAGCGCGGAGCAGCGACCTGAGCAGCAGGCAAGAT	166
QY			
DB	41	AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys	60
QY			
DB	167	GCTGAATAATGCTGAGGCAAAAGCTCAGGGGCTCCCGGGGAGCTTGTGACATCGCGTGC	226
QY			
DB	61	LysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp	80
QY			
DB	227	AAGGTGTCCAGGCTTACCTGGGGCAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT	286
QY			
DB	81	AlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrTyrSerLeu	100
QY			
DB	287	GCCGTGGAGGACCTCACTGAGCCGAGTGGAGGACCTGACCCAGCAGTACTTCCCTC	346
QY			
DB	101	ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu	120
QY			
DB	347	GTTC-----TGCCAGGCTCTG	363
QY			
DB	121	LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrp	140
QY			
DB	364	CTGAATAGAAATCAGCTGTGAAACCTTCAGACGGACATTGACGGGCTCCGGAAACATCTGG	423
QY			
DB	141	IleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu	160
QY			
DB	424	ATTATAAGCCCGCGGCCAAGTCCCGGGCCGAGACATAGTGTGATGGACCGTGTGGAG	483
QY			
DB	161	GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValVal	180
QY			
DB	484	GAGATCCTGGAGCTGGCAGCTGCAGACCACTCTTTCCAGGGACAAACAAGTGGGTGTC	543
QY			
DB	181	GlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrp	200
QY			
DB	544	CAGAAGTATACGAGACCGCTGCTCATCTGTGTGACCAACCAAGTTCGACATCAGACAGTGG	603
QY			
DB	201	PheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArg	220
QY			
DB	604	TTCTCTGTACGAGCTGGAAACCCCTTGACCATCTGGTTCTACAAGGAGAGTACTTTCGGG	663
QY			
DB	221	PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn	240
QY			
DB	664	TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGGCCCATCCACTGTGTGCAACAAC	723
QY			
DB	241	AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn	260
QY			
DB	724	GCCGTCCAGAGTACTGGAAGATGATGTGGCCGCGAGCCCTGTGTCCCGCAGACAAAC	783
QY			
DB	261	MetTrpThrSerThrArgPheGlnGlnLysLeuGlnArgGlnGlyArgGlyAlaValTrp	280
QY			
DB	784	ATGTGGACAGCAGCAGGTTCAGGAGTACCTGCAGCGCAGCGCGCGTGGCGGTGG	843
QY			
DB	281	GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln	300
QY			
DB	844	GGCAGCGTCATCTACCCGTCATGAAGAGGCCATCGCCACGCCATGAAGTGGGCCAG	903
QY			
DB	301	AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly	320
QY			
DB	904	GACCAGTGGAGCTCCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCTTGGG	963
QY			
DB	321	ArgAspPheArgProTrpLeuLeuGluLeuLeuTrpArgGlnProValValGluProPro	380
QY			
DB	964	AGGGACTTCAGGCCCTGGCTGATCGAGATCAATTCAGGCCCCACCATGCACCGTCCACG	1023
QY			
DB	341	ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp	360
QY			
DB	1024	CCGGTCAGGCCCGAGCTGTGTGCAGAGGTGCAGGAGGACACCATCAAGTGGCGGTGGAC	1083
QY			
DB	361	ArgSerCysAspIleGlyAsnPheGluLeuLeuLeuTrpArgGlnProValValGluProPro	380
QY			
DB	1084	CGCAGCTGTGACATCGGCAACTTCGAGCTCTCTGTGGAG-----CCGGTGTGTGAGCCGCC	1139
QY			
DB	381	ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaArgArgGln	400
QY			
DB	1140	CCATTTCAGCGGGTCCGACCTCTCGCTGGCGGGCGCTCAGTGTGAGGAGAGCCAGGAGCAG	1199
QY			
DB	401	ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys	420
QY			
DB	1200	GTGCTGTCCCGCTGTGCAACCTCAAGGCCTCGGCTCGCTGTGTGGACGCGCAGCGCTGAAG	1259
QY			
DB	421	AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln	440
QY			
DB	1260	GCACGGGGCCCTCGGCCATGCCAGACCTCGCCAGGGAGACCCCATCACAGCTCTCCAG	1319
QY			
DB	441	ArgAspLeuGlyLeuLysGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuArg	460
QY			
DB	1320	CGGGACTTGGGACTGAGAGGAGAGAGGGGCTCCCTCTGGCTTCTGTGGACCCCTTAAGG	1379
QY			
DB	461	GlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal	480
QY			
DB	1380	GGGGCAGCGAGAGCGGTGGAGCCGACAGCCGACCCGACCAAGCTGTGGGAAGGTG	1439
QY			
DB	481	GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro	500
QY			
DB	1440	GAGTTCGGGGCTCGCCCTGTGCGCACGTGGACAGTGGGGCCCAACACCGGTGTCCCC	1499
QY			
DB	501	ValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGluPro	520
QY			
DB	1500	GTAGCCAGCCCGCCAAAGCTGGATCCAAACACAGCTAAATGAGCACCCCGCTGGAGCT	1559
QY			
DB	521	ValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGlyLysGly	540
QY			
DB	1560	GTGCTGCGGAGCTGAACAGACAGAGGCGCGTGTGCTCGCGCCCGCGGAGGAAAGT	1619
QY			
DB	541	Ser	541
QY			
DB	1620	TCA	1622
QY			
DB			

RESULT 4

ADJ93363
ID ADJ93363 standard; DNA; 3465 BP.
XX
AC ADJ93363;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human BGS-42 protein-related DNA clone C SeqID11.
XX
KW testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nontropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary cancer; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; ds.
XX
OS Homo sapiens.
XX
XX W02004005487-A2.
XX
XX 15-JAN-2004.
XX
XX 09-JUL-2003; 2003WO-US021605.
XX
XX 09-JUL-2002; 2002US-0394725P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Feder JN, Wu S, Nelson TC;
XX
XX WPI; 2004-099381/10.
XX
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
XX useful for preventing, treating or ameliorating a medical condition, e.g.
XX PT aberrant cellular proliferation, reproductive disorders or testicular
XX disorders.
XX
XX Example 4; SEQ ID NO 11; 343pp; English.
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
XX -like polypeptide, designated the BGS-42 polypeptide. The invention may
XX be useful for the development of compounds with a cytosolic, respiratory
XX -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
XX antiinflammatory, anabolic, hypertensive, osteopathic, nontropic,
XX antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
XX immunosuppressive, antiseborrheic or dermatological activity acting as
XX tyrosine ligase modulators. In addition, the disclosed sequences may be
XX useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
XX used for diagnosing a pathological condition or a susceptibility to a
XX pathological condition in a subject, and for preventing, treating or
XX ameliorating a medical condition, such as a disorder related to aberrant
XX tubulin ligase activity, a disorder related to aberrant tubulin-
XX carboxypeptidase activity, aberrant cellular proliferation, reproductive
XX disorders, testicular disorders, testicular cancer, pulmonary disorders,
XX lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
XX neural disorders, brain cancer, liver cancer, or proliferative condition
XX of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
XX polypeptide, polynucleotide, or their modulators are also useful for
XX treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
XX disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
XX -42 polypeptide can be used as a preventive agent for immunological
XX disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
XX disease or scleroderma. The antibodies may be used to purify, detect and
XX target the BGS-42 polypeptides. The present sequence is that of a DNA
XX clone sequence which is related to the invention.

XX
SQ Sequence 3465 BP; 667 A; 1074 C; 1126 G; 598 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8,51e-166 Length: 3465
Score: 2693.00 Matches: 516
Percent Similarity: 95.4% Conservatives: 0
Best Local Similarity: 95.4% Mismatches: 3
Query Match: 94.4% Indels: 23
DB: 12 Gaps: 1

US-10-635-977-2 (1-541) x ADJ93363 (1-3465)

QY 1 MetAlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerArg 20
Db 1678 ATGGCATCCAGCATCTCAAGTGGTGGTGCAGCCACGAGTGCAGCAGGACGACAGA 1737

QY 21 SerLysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerArgGlnAsp 40
Db 1738 AGCAAGCCCGAGGACGAGAGGAGGAGCGCGGAGCAGCGACCTGAGCAGCAGGCAAGAT 1797

QY 41 AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys 60
Db 1798 GCTGAAAATGCTGAGCAAGCTCAGGGGCTCCCGGGGAGCTTGTGCACATCGCGTGC 1857

QY 61 LysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp 80
Db 1858 AAGTGTGTCCAGGCTTACCTGGGCGAGCTGGAGCATGAGGACATCGACAGTCAGCAGAT 1917

QY 81 AlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrTyrSerLeu 100
Db 1918 GCCGTGGAGACCTCACTGAGGCCGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC 1977

QY 101 ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu 120
Db 1978 GTTCATGGCGATGCTTTTCATCTCCAATTCAAGAAATTACTTTTCGAGTCCAGGCTCG 2037

QY 121 LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrp 140
Db 2038 CTGAATAGATACGCTGTGAACCTTCAGCGGACATTTGACGGGCTCCGGACATCTGG 2097

QY 141 IleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
Db 2098 ATTATAAGCCCGCGCCCAAGTCCCGGGCGG----- 2129

QY 161 GluIleLeuGluLeuAlaAlaAlaAspHisProLeuSerArgAspAsnLysTrpValVal 180
Db 2130 -----AGGGACCAACAAGTGGTGGTGC 2150

QY 181 GlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrp 200
Db 2151 CAGAAGTACATCAGACAGCGCGCTGCTCATCTGTGACACCAAGTTCCGACATCAGACAGTGG 2210

QY 201 PheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArg 220
Db 2211 TTCTCGTCAAGGCTGGAACCCCTGACCATCTGGTTCTACAAAGGAGAGTACTTTCGGG 2270

QY 221 PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn 240
Db 2271 TTCTCAACTCAGCGCTTCTCCCTGGACAAAGCTGGACAGCGCCATCCACCTGTGCAACAC 2330

QY 241 AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn 260
Db 2331 GCGCTCCAGAAGTACCTGAAGAATGATGTGGCGCGACGCCCTCTGCTCCCGCAGCACACAC 2390

QY 261 MetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTrp 280
Db 2391 ATGTGGACACGACCCAGTTCAGGAGTACTCTGACGCGCCAGGCGCGCTGGCGCGTGTGG 2450

QY 281 GlySerValIleTyrProSerMetLysAlaIleAlaHisAlaMetLysValAlaGln 300
Db 2451 GGCAGCGCTCATACCGTCCATCAGAGAGGCCATCGCCACGCCCATGAAGGTGGCCCGAG 2510

QY	41	AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys	60	QY	401	ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys	420
Db	273	GCTGAAATGCTGAGCAAGCTCAGGGGCTCCGGGGCAGCTTGTGGACATCGCGTGC	332	Db	1286	GTGTCGCCCTGTCGAACCTCAAGGCTCGGCTCGCTGTTGGACGCGCAGCGCTGAAG	1345
QY	61	LysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp	80	QY	421	AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln	440
Db	333	AAGGTGTCCAGGCCCTACCTGGGGCAGCTGGACATGAGACATCGACAGTCAGCAGAT	392	Db	1346	GCACGGGGCCCTCGGCCATGCCAGACCTGCGCCAGGAGCCCCCATCACAGCTCTCCAG	1405
QY	81	AlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrTyrSerLeu	100	QY	441	ArgAspLeuGlyLeuLysGlnLysGlyLeuProLeuAlaLeuLeuAlaProLeuArg	460
Db	393	GCGCTGGAGACCTCACTGAGGCCGAGTGGAGGACCTGACCCAGCAGTACTACCTC	452	Db	1406	CGGGACTTGGGACTGAAGGAAGAGAGGGGCTCCCCCTGGCTTGTGGACCTTTAAGG	1465
QY	101	ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu	120	QY	461	GlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal	480
Db	453	GTTTCATGGCGATGCTTTCATCTCCAATTCAAGAAATTACTTTTCGAGTGCCAGGCTCG	512	Db	1466	GGGGCAGCCGAGAGCGGTGGAGCCGCACAGCCACCCGACCAAGCTGCTGGGAAGGTG	1525
QY	121	LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTyr	140	QY	481	GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro	500
Db	513	CTGAATAGATCACTGCTGTGAACCTTCAGACGGACATTTACCGGGCTCCGGAACATCTGG	572	Db	1526	GAGCTCCGGCTGCTCCCTGTCGCCACGTGGACAGTCAGGCCCAAAACACCGGTGCC	1585
QY	141	IleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu	160	QY	501	ValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGluPro	520
Db	573	ATTATAAGCCCGCGCAAGTCCGGGGCG-----	604	Db	1586	GTAGCCAGCCGCCCAAAAGCTGGGATCCAAACACAGCTAAATGAGCACCGCTGGAGCCT	1645
QY	161	GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValVal	180	QY	521	ValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGlyLysGly	540
Db	605	-----AGGACAAACAAGTGGTGGTC	625	Db	1646	GTGCTCGGAGCTGAAGACAGCAGAGGGCGCTGCTCCGCCCGCGAGGAAAGGT	1705
QY	181	GlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrp	200	QY	541	Ser 541	
Db	626	CAGAAGTACATCAGACGCGCTGCTCATCTGTGACACCAAGTTCCGACATCAGACAGTGG	685	Db	1706	TCA 1708	
QY	201	PheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArg	220	RESULT 6			
Db	686	TTCTCTGTCAGGACTGAACCCCTGACCATCTGGTTCTACAGGAGAGTTACTTGGCG	745	ID	ADQ17814/c		
QY	221	PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn	240	AC	ADQ17814;		
Db	746	TTCTCAACTCAGGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCCAACAC	805	XX			
QY	241	AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn	260	DT			
Db	806	ACCGTCCAGAAGTACCTGAAGAATGATGGCGCGACGCCCTGCTGCGCCGCACACAC	865	DE	26-AUG-2004 (first entry)		
QY	261	MetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTrp	280	XX	Human soft tissue sarcoma-upregulated DNA - SEQ ID 631.		
Db	866	ATGTGGACACAGCAGGTTCCAGGAGTACTTCAGCGCCAGGGCGTGGCGCGTGTGG	925	KW	soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;		
QY	281	GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln	300	KW	ds.		
Db	926	GGCAGCGTCATCTACCGCTCATGAAGAGGCCATCGCCACGCCATGAAGGTGGCCACG	985	OS	Homo sapiens.		
QY	301	AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly	320	XX	WO2004048938-A2.		
Db	986	GACCACGTGAGCGCTCCCAAGACAGCTTTGAGCTCTACGGGGCTGACTTCGTCCTTGGG	1045	PN	10-JUN-2004.		
QY	321	ArgAspPheArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThr	340	PD			
Db	1046	AGGACTTCAGGCCCTGGCTGATCGAGATCAATTCAGGCCACCATGACCCGCTCCACG	1105	XX	26-NOV-2003; 2003WO-US038193.		
QY	341	ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp	360	PF	26-NOV-2002; 2002US-0429739P.		
Db	1106	CCGGTCACGGCCCGAGCTGTGTCACAGGTGCGAGGAGCACCATCAAGGTGGCGTGGAC	1165	PR	(PROT-) PROTEIN DESIGN LABS INC.		
QY	361	ArgSerCysAspIleGlyAsnPheGluLeuTrpArgGlnProValValGluProPro	380	XX			
Db	1166	CGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGGAGGCGCGGTGGTGGAGCCGCC	1225	PI	Aziz N, Ginsburg WM, Zlotnik A;		
QY	381	ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaArgArgGln	400	XX	WPI; 2004-441208/41.		
Db	1226	CCATTACGGGGTCCGACCTCTGGTGGCGGGCTCAGTGTGAGGAGAGCCAGGAGCGAG	1285	PT	Early detection of soft tissue sarcoma comprises determining expression		
				PT	of a gene in a first soft tissue sample and a normal soft tissue sample		
				PT	and comparing the gene expression, also useful in treating soft tissue		
				PT	sarcoma.		
				XX	Example 2; SEQ ID NO 631; 210pp; English.		
				PS			
				XX	The invention relates to a novel method for detecting soft tissue sarcoma		
				CC	which comprises obtaining a first soft tissue sample from an individual,		
				CC	and a normal soft tissue sample from the same or different individual,		
				CC	determining the expression of a gene in both samples and comparing the		
				CC	expression of the gene in both soft tissue samples, where a higher level		

CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cytostatic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC DNA of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.
 XX

SQ Sequence 101270 BP; 24151 A; 28044 C; 28100 G; 20975 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 4,62e-106 Length: 101270
 Score: 1802.00 Matches: 438
 Percent Similarity: 32.2% Conservative: 1
 Best Local Similarity: 32.1% Mismatches: 3
 Query Match: 63.1% Indels: 923
 DB: 12 Gaps: 3

US-10-635-977-2 (1-541) x ADQ17814 (1-101270)

QY 100 LeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyPheSerGlnCysGlnAla 119
 DB 47222 CTTTGTAGTGGCGATGCTTTTCATCTCCATTCAGAAATATCTTTTGGAGTGCAGGCT 47163
 QY 120 LeuLeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIle 139
 DB 47162 CTGCTGAATPAGAATCACGCTGTGAACCTCAGACGGACATTCAGCGGCTCGGAACATC 47103
 QY 140 TrpIleIleLysProAlaAlaLysSerArgGlyArg----- 151
 DB 47102 TGGATTATTAAGCCCGCGCCCAAGTCCCGGGGCGG-AGGTGAGTCCCTGGTGGCGATGC 47044
 QY 151 ----- 151
 DB 47043 TCCTTGACCGGTTTCATAGCCACCACCGGCCACACTGTGATCCAGGAGCCGTGAGTGA 46984
 QY 151 ----- 151
 DB 46983 GGACAGTTTCAGGCCACAGACCACAGGAAGGAAGGAGGAGCAGAAACGCCGAGCA 46924
 QY 151 ----- 151
 DB 46923 GCGTGGGGTTCAGGACTCAGCGCTGGAGCAGAGCTGAGGCGTCTTCACAGCATGTGT 46864
 QY 151 ----- 151
 DB 46863 TCCGATCACGGCAGGATCCGGTTGATGGTTAATCAGGATTTCTCTCTTTGATTCA 46804
 QY 151 ----- 151
 DB 46803 CAAACTGCTATGCTTTCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 46744
 QY 151 ----- 151
 DB 46743 TGAACGAGTTTTCATCTTGTGTGCCAGGCTGAAGTGAAGTGAAGTGAAGTGAAGTGA 46684
 QY 151 ----- 151
 DB 46683 TGCACCTCCACCTTCAGGTTTAAGCAATTTCTCTGCTTAGCCTCCCAAGTAGCTGGG 46624
 QY 151 ----- 151
 DB 46623 ACTACAGCGGCCACACACCGGCTAATTTTGTATTTTATTTTATTTTATTTTATTTT 46564
 QY 151 ----- 151
 DB 46563 CCATGTTGGCCAGGCTGCTTGTAACTCCCGACCTCAAGTGATTCACCTGCTCGGCTC 46504
 QY 151 ----- 151
 DB 46503 CCAAGTCTGGGATTACAGGATTAAGCCACATGCCCATCTCTCTGCTCCCTTTTAA 46444

QY 151 ----- 151
 DB 46443 AGTGTCTCTGTCCCTGCCCTTCACTGATCTCTGTGTCCCTCTGGGTCTCAGAGGCTG 46384
 QY 151 ----- 151
 DB 46383 GTTCTGTTGGGGAGAGCCTGGGGGACTCAGGCACTGGGCTTTAGCCTCATCATGAGCGC 46324
 QY 151 ----- 151
 DB 46323 TAAGCTTCCCTGCTCCCTCCATCGCCCCCACCACCAACCATGCTGCTGCTGGCCTTGGACAA 46264
 QY 151 ----- 151
 DB 46263 AGCCAGTGCAGGTCTCAGSGCCAAGTGCAGCCCTGCCCCAGGGCTCAGGAGCTCAGGTC 46204
 QY 152 -----AspIleValCys 155
 DB 46203 CCGCTGCACAGTGGCTGTGCATGTGGTGTGAGCGGTGGCCTCTTTCAGACATAGTGTGC 46144
 QY 156 MetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAsp 175
 DB 46143 ATGGACCGTGTGGAGAGATCTCGAGCTGGCAGCTGCAGACCCCTCTTTCAGGAC 46084
 QY 176 AsnLysTrpValValGlnLysTyIleGluThrProLeuLeuIleCysAspThrLysPhe 195
 DB 46083 AACAAAGTGGTGGTCCAGAAGTACATCGAGACGCCCTGCTCATCTGTGACACCAAGTTC 46024
 QY 196 AspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyIys 215
 DB 46023 GACATCAGACAGTGGTCTCTCGTCAAGGACTGGAAACCCCTCACCATCTGTGTCTACAA 45964
 QY 216 GluSerTyLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAsp----- 232
 DB 45963 GAGATTACTTGGCTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGGTCAGTG 45904
 QY 232 ----- 232
 DB 45903 CGGTGGGCGACTGGACACCTCGGGCCAGGNAATGGGCTGTGCTGCTGCTGCGAGCA 45844
 QY 232 ----- 232
 DB 45843 GCCGAGTGAAGAGCAGCACGCGGCGAGGCGCCCTCACTGCCACCTGCCACACGAGAGC 45784
 QY 232 ----- 232
 DB 45783 ACAGATGGGGAACCTGAGGCTTGGAGGCTGATGATCTCCCGAGGAGCAGGCCCAGGCC 45724
 QY 232 ----- 232
 DB 45723 CACCGGCTTCAAAGCCAGCGCCCTACACCCACGCTCTGTGCATCATCTCTGGAAATC 45664
 QY 232 ----- 232
 DB 45663 TTAAGGSCACCAATTTTGTCTGTGCCAGGGGATTTGAATGGTGTGGGAGCAGCCTGGG 45604
 QY 232 ----- 232
 DB 45603 CCCTGGGCGCAGGGGAGGCGGCTGTGGGTGCTGTGGAGGCTCAGATATCCGCAAGTG 45544
 QY 232 ----- 232
 DB 45543 TGCTGCCAGGCTGATGGACCCCACTCACCCGGGTCCCGCTGGGAGGAGGCGCA 45484
 QY 232 ----- 232
 DB 45483 GAGCCTGTGGAGTCTCAGGAGTGTGTGTGTCACACAGCAGCGCTGCCACCTCTGTCT 45424
 QY 232 ----- 232
 DB 45423 GCACCAAGGCTTCTCTGACTTATTTGTGGCCCTTGTGTGGTGTGAGCGCTAGGGGACC 45364
 QY 232 ----- 232

Db	45363	TGACGGCTGGGACAGGAAGGGGGTCTGCTTTTACGAGTGAGAAATCCAGGAGGCCTGAGA	45304
QY	232	-----	232
Db	45303	GTACCGGACACAGACCTTTCAGGGAGTTAACTCCGGCCACGAGCGGTGCGGTCA	45244
QY	232	-----	232
Db	45243	TGCCTGGTGGGTCAATGCTCGAAGTCACACCAGAAGAGGCTGGAGGAGCTGGCCAGGG	45184
QY	232	-----	232
Db	45183	GCTACGACCCTCACCTGCTCACTGCTCTGCTCTGGTCAAGTGGGCCCTGGAGGGGGTCCAGGAAC	45124
QY	232	-----	232
Db	45123	TCACCTGTGCAAAAGCAGCTCCACGAGCGGCAGTGAGAAGCCAGGCCCCCTGCCAGCG	45064
QY	232	-----	232
Db	45063	CCCAITTTGGGCTCCAGCTGTGAGCTGTGCGCTGGCCAAATCCCATGGSCCTCTCTGAG	45004
QY	232	-----	232
Db	45003	CCCTGTTTCTCCATCTGGGGACAGTAACGCTCATCCACAGTGGGGCCCTCCAGCAGAGGGT	44944
QY	232	-----	232
Db	44943	TGGAGTCCCTCAGGGGCGAGGGTTCTTAGACCAGCTGGGTGGGCTCAGCACGGCTGGAG	44884
QY	232	-----	232
Db	44883	GACCGGCTTTGGTGGCTCTCTGGGCCAAGGAGGCTGAGCACCTGCCCCCTGCTCCCCC	44824
QY	233	---SerAlaIleHisLeuCysAsnAsnAlaValGlnIlystYrLeuLysAsnAspValGly	251
Db	44823	TGACGGCCCATCACCTGTGCAACACGCGCTCAGAAGTACCTGAAGAATGATGGGC	44764
QY	252	ArgSerProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeu	271
Db	44763	CGCAGCCCCCTGTGCCCGCACACAACATGTGCACGACACAGGTTCACAGGAGTACCTG	44704
QY	272	GlnArgGlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAla	291
Db	44703	CAGCGCCAGGCGCTGGCGCGTGTGGGCGAGCGTCACTACCCGCTCCATGAAGAAGGCC	44644
QY	292	IleAlaHisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGlu	311
Db	44643	ATCGCCACGCCATGAAGGTGGCCAGGACACCTGGAGGCTCGCAAGAACAGCTTTGAG	44584
QY	312	LeuTyrGlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluIleAsn	331
Db	44583	CTCTACGGGCTGACTTCGTCTTGGGAGGAGCTTCAGGCCCTGGCTGATCGAGATCAAT	44524
QY	332	SerSerProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGln	351
Db	44523	TCCAGCCCCCATGCACCCGCTCACCGCGTTCACGGCCCGAGCTGTGTGCACAGGTGCAG	44464
QY	352	GluAspThrIleLysValAlaValAspArgSerCysAspIleGlyAsnPheGluLeuLeu	371
Db	44463	GAGACACCATCAAGGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTG	44404
QY	372	TrpArg-----	373
Db	44403	TGGAGGCGAGGTGAGCCACGCCGCCCTCTGGGACTTTTGGTGGCCGCCCGCATGTGGATG	44344
QY	373	-----	373
Db	44343	CTTTGCCACCTGGTCACTGGGCGAGGGTGGAGCTGCAGGCGCGCAGTGCAGCGGGGTCTA	44284
QY	373	-----	373
Db	44283	TGGAGSACACCGATCCTGCTCTGTGTGACCATGTCTGGTCCGCTCCGCTAGGGGCCAGGCCTGCC	44224
QY	373	-----	373
Db	44223	CTGCTCTCCATGGGGCTCAGTGCCTTCGTTCGGACCCGAGAAAGGGGCTACTCCAGGGGTG	44164
QY	373	-----	373
Db	44163	CGGGGAGGGGCGCGCTCACTCAGGCTCAGTGCCTCTGTCCAGCTTTCTCCTCCTCTTG	44104
QY	373	-----	373
Db	44103	TTCCACAGCTCACCCAGTCTGACTGTGTGTAGAGTGGGGCTCCATGCCCGCAGCCTC	44044
QY	373	-----	373
Db	44043	CAGCCACCATATGCTGCTCTTTTGGCTTGGCTTGCAGTGCCTCCTCCACTCTGAGGATCT	43984
QY	373	-----	373
Db	43983	TTGCGGGGCTCCCTGTGCTGCGGATGTCTGGAGTCATTGGGAAGCTGGCCAATGCC	43924
QY	373	-----	373
Db	43923	TCTCTCTTGGGCTCTCTGTGGCTAACTCGGGAAGAGCAGCTGAGTCCCTGGGTGT	43864
QY	373	-----	373
Db	43863	GGGGCCCCCAGGCTGACCCCCCAAGCCTCAGGAAGCCTTGGTTCCCCGGCAGGTTCCAG	43804
QY	373	-----	373
Db	43803	AGCCTGGGGCTTCTGGGGGTCTGCTGGGACCCAGGTTTCACTGTTCACTCCTCTGAG	43744
QY	373	-----	373
Db	43743	TCAGTGGGGTCTGTGCCCCGACAGCCCTCTCTGCACAGACCCCTCTGCAGGGACCC	43684
QY	374	-----GlnProValValGln	378
Db	43683	CCCCATCCCGGCACAGCCCCACACCCCAACCTGTTCCTTCCCACAGCGGTGGTTGA	43624
QY	378	uProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgArgAlaAr	398
Db	43623	GCCGCCCCCATTCAGCGGTCCGACCTCTGCGTGGGGGCTCAGTGTGAGGAGGCCAG	43564
QY	398	gArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnPr	418
Db	43563	GAGGCAGGTGCTGCCGCTCTGCAACCTCAAGGCCTCGGCTCTGTCGACGCGCAGCC	43504
QY	418	oLeuLysAlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAl	438
Db	43503	GCTGAAGGCAACGGGGCCCTCGGCCATGCGCACCCCTGCGGAGGCCATCCACGAGC	43444
QY	438	aLeuGlnArgAspLeuGlyLeuLysGluGlyLysGlyLeuProLeuAlaLeuLeuAlaPr	458
Db	43443	CTTCCAGCGGACTTGGGACTGAAGNAGAGAGGGGCTCCCCCTGGCTTGTGGGCACC	43384
QY	458	oLeuArgGlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGl	478
Db	43383	CTTAAGGGGGCAGCGAGAGCGGTGGAGCCGCACAGCCCCACCGCACCAAAAGCTGTGG	43324
QY	478	YLysValGluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGl	498
Db	43323	GAAGGTGGAGCTCCCGGCTGTCGCCAGTGGAGCAGTCCAGCCCCAACACCCGG	43264
QY	498	yValProValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLe	518
Db	43263	TGCCCCGTAGCCAGCCCCCAAAAGCTGGATCCAAACAGACTAAATGGCACCCGCT	43204
QY	518	uGluProValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGl	538
Db	43203	GGAGCCTGTCTGCGGGCTTGAAGACAGCAGAGGGGCGCTGCTGCCGCCGGAGG	43144

QY 538 ylysGlySer 541
 Db 43143 AAAAGGTAAAC 43134
 RESULT 7
 ID ADJ93359 standard; cDNA; 726 BP.
 XX ADJ93359;
 XX 06-MAY-2004 (first entry)
 XX Human BGS-42 cDNA sequence SeqID3.
 XX testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
 KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nontropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
 KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human; gene; ss.
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH CDS 1..726
 FT /*tag= a
 FT /product= "Human BGS-42 protein"
 FT /partial
 FT /note= "No start or stop codon"
 XX
 PN WO2004005487-A2.
 XX
 PD 15-JAN-2004.
 XX
 PF 09-JUL-2003; 2003WO-US021605.
 XX
 PR 09-JUL-2002; 2002US-0394725P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Feder JN, Wu S, Nelson TC;
 XX
 DR WPI; 2004-099381/10.
 XX
 PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
 PT useful for preventing, creating or ameliorating a medical condition, e.g.
 PT aberrant cellular proliferation, reproductive disorders or testicular
 PT disorders.
 XX
 PS Disclosure; SEQ ID NO 3; 343pp; English.
 XX
 CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
 CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytostatic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, nontropic,
 CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 CC immunosuppressive, antiseborrheic or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC used for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant
 CC tubulin ligase activity, a disorder related to aberrant tubulin-

CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of a partial
 CC cDNA which encodes the human BGS-42 protein of the invention.
 XX

SQ Sequence 726 BP; 157 A; 227 C; 214 G; 128 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8,81e-76 Length: 726
 Score: 1299.00 Matches: 242
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 45.5% Indels: 0
 DB: 12 Gaps: 0

US-10-635-977-2 (1-541) x ADJ93359 (1-726)

QY 133 lleAspGlyLeuArgAsnIleTptllelleLysProAlaAlaLysSerArgGlyArgAsp 152
 Db 1 ATTGACGGGCTCCGGAAACATCTGGATTATTAAGCCCGCGGCCAAGTCCCGGGCCGAGAC 60
 QY 153 lleValCysMetAspArgValGluGluLeuGluLeuAlaAlaAspHisProLeu 172
 Db 61 ATAGTGTGCATGGACCGGTGTGGAGGAGATCCTGGAGCTGGCAGTGCACACCCCTCTT 120
 QY 173 SerArgAspAsnLysTptValGlnLysTyrIleGluThrProLeuLeuLeuCysAsp 192
 Db 121 TCCAGGGACAAACAGTGGGTGTCCAGAAAGTATACGAGACCCCGCTCTCATCTGTGAC 180
 QY 193 ThrLysPheAspIleArgGlnTptPheLeuValThrAspTptAsnProLeuThrIleTpt 212
 Db 181 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGGACTGGAAACCCCTGACCATCTGG 240
 QY 213 PheTyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAsp 232
 Db 241 TTCTACAAGGAGAGTACTTTCGGTTCCTCAACTCAGCGCTTCCTCCCTCGACAAGCTGGAC 300
 QY 233 SerAlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArg 252
 Db 301 AGCGCCATCCACCTGTGTCAACCAACCGCGTCCAGAAAGTACCTGAAGAATGATGTGGCCGC 360
 QY 253 SerProLeuLeuProAlaHisAsnMetTptThrSerThrArgPheGlnGluTyrLeuGln 272
 Db 361 AGCCCCCTGCTGCCCGCACACACATGTGGACACGACACAGTTCCAGAGGTACTCTGGAC 420
 QY 273 ArgGlnGlyArgGlyAlaValTptGlySerValIleTyrProSerMetLysLysAlaIle 292
 Db 421 CGCCAGGCGCGTGGCGCGTGTGGGGCAGCGTCTATCCCGTCCATGAAGAAGGCCATC 480
 QY 293 AlaHisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeu 312
 Db 481 GCCCAGCGCATGAAGGTGGCCCGCAGGACCCAGCGCTCGCAGAACAGCTTTTGGAGTTC 540
 QY 313 TyrGlyAlaAspPheValLeuGlyArgAspPheArgProTptLeuIleGluLeuAsnSer 332
 Db 541 TACGGGGCTGACTTCGTGCTTGGGAGGAGACTTCAGGCCCTGCTGCTATCGAGATCAATTCC 600
 QY 333 SerProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGlu 352
 Db 601 AGCCCCCATGCACCCCGTCCACCGCGTCAACCGCCAGCTGTGTGTCACAGGTGTCAGGAG 660
 QY 353 AspThrIleLysValAlaValAspArgSerCysAspIleGlyAsnPheGluLeuLeuTpt 372
 Db 661 GACACCATCAAGGTGGCGCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGG 720

QY 373 ArgGln 374
Db 721 AGGCAG 726

RESULT 8
AAS99894
ID AAS99894 standard; cDNA; 2380 BP.
XX AC
XX AAS99894;
XX 12-MAR-2002 (first entry)
DT
XX Polynucleotide encoding human cytoskeleton-associated protein (CYSKP) #5.
DE
XX Human; cytoskeleton-associated protein; CYSKP; autoimmune disorder; ss;
KW cell proliferative disorder; inflammatory disorder; prion disease;
KW vesicle trafficking disorder; gastrointestinal disorder; muscle disorder;
KW neurological disorder; cell motility disorder; reproductive disorder;
KW spinal cord disease; central nervous system disorder; mental disorder;
KW gene therapy; cancer.
XX
OS Homo sapiens.
XX
XX WO200185942-A2.
XX
XX 15-NOV-2001.
XX
XX 03-MAY-2001; 2001WO-US014355.
XX
XX 05-MAY-2000; 2000US-0201960P.
XX 08-MAY-2000; 2000US-0202729P.
XX 05-JUN-2000; 2000US-0209705P.
XX 07-JUN-2000; 2000US-0210149P.
XX 21-JUN-2000; 2000US-0213215P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Yue H, Tang YT, Au-Young J, Lu DAM, Baughn MR, Hillman JL;
PI Azimzai Y, Lal P, Yao MG, Bandman O, Burford N, Batra S, Kearney L;
PI Policky JL;
XX
XX WFI; 2002-062248/08.
XX P-PSDB; AAU74334.
XX
XX New cytoskeleton-associated proteins and polynucleotides, useful for
PT diagnosing, preventing and treating cell proliferative, autoimmune,
PT inflammatory, neurological, cell motility, reproductive and muscle
PT disorders.
XX
XX Claim 5; Page 171; 194pp; English.
XX
XX The invention relates to human cytoskeleton-associated polypeptides
CC (CYSKP) and their associated polynucleotide sequences. The sequences are
CC useful in the treatment of disorders associated with overexpression or
CC underexpression of CYSKP in a patient. The disorders include cell
CC proliferative disorders (such as cancer, actinic keratosis,
CC arteriosclerosis, cirrhosis, hepatitis and psoriasis),
CC autoimmune/inflammatory disorders (such as, asthma, atherosclerosis,
CC osteoporosis, Crohn's disease, rheumatoid arthritis, diabetes mellitus
CC and anaemia), vesicle trafficking disorders (such as
CC hypercholesterolaemia, diabetes insipidus, Grave's disease and goitre),
CC gastrointestinal disorders, prion diseases, neurological disorders (such
CC as epilepsy, stroke, cerebral neoplasms, Alzheimer's disease,
CC Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis
CC and other motor neuron disorders), cell motility disorders, reproductive
CC disorders (such as endometriosis and polycystic ovary syndrome), muscle
CC disorders (such as myocarditis, migraine, hypertension, hypoglycaemia,
CC myocardial infarction, epilepsy and muscular dystrophy), spinal cord
CC diseases, central nervous system disorders (such as Down syndrome and
CC cerebral palsy) and mental disorders (such as anxiety and schizophrenia).
CC Sequences AAS99890-AAS99923 represent cDNA molecules encoding human CYSKP
CC of the invention

XX
SQ Sequence 2380 BP; 580 A; 685 C; 670 G; 445 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2.67e-60 Length: 2380
Score: 1070.50 Matches: 250
Percent Similarity: 54.8% Conservative: 81
Best Local Similarity: 41.4% Mismatches: 198
Query Match: 37.5% Indels: 76
DB: Gaps: 12

US-10-635-977-2 (1-541) x AAS99894 (1-2380)

QY 2 AlaSerSerIleuLeuYsrPValValSerHisGlnSerCysSerArgSerArgSer 21
Db 285 GCCCGCAAGCTTCTCAAGCTGGTGGTGAAGTCTGAG-----TGGAGTCA 329
QY 22 LysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerArgGlnAspAla 41
Db 330 TACCCTATTTCAGGCAGTACAGGAAGAGGCTCAGGAGCAAGCAGCCCAAGAACAGGAG 389
QY 42 GluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLys 61
Db 390 AAAAACCCAGTGTG-----GTGTCCCGCAGAGTTGTGGATGAAGCTGTGTGT 437
QY 62 ValCysGlnAlaTyrluGlyGlnLeuGluHisGluAspIleAspThrSerAlaAspAla 81
Db 438 GCGTGGGAGGAGTACCTTTAGCAACTTGGCCCACTGACATCGAACAGGACCTGGAGGCC 497
QY 82 ValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrluSerLeuVal 101
Db 498 CGCGTGTACTACACCCCGAGGGCTGGTCCCTCTTCTCCAGCGCTACTACCAAGTGTTC 557
QY 102 HisGlyAspAlaPheIleSerAsnSerArgAsnTyrluPheSerGlnCysGlnAlaLeuLeu 121
Db 558 CACGAAGGGGCGAAGCTCAGGCACTCGACCTCAGGCTCCAGCGCTGTGAGGACATCTCTG 617
QY 122 AsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIle 141
Db 618 CACGACGTGAGCGCGCTGGTGTACCCAGATAGATGAGAGGGATCGCAACATCTGGATC 677
QY 142 IleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGlu 161
Db 678 GTCAAGCCAGGAGCCCAAGTCCCTGGACGAGGATCATGTGATGGACCACTTGGAGGAG 737
QY 162 IleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnIleTyrluValGln 181
Db 738 ATGCTGAAGCTGGTGAACCGCAACCCCGTGGTGTATGAAGAGCGCAAGTGGTGTGCAG 797
QY 182 LysTyrlleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPhe 201
Db 798 ARGTTATATTGAGCGGCGCCCTCCCTCATCTTTGGGCCCAAGTTTGACCTCAGACAGTGTTC 857
QY 202 LeuValThrAspTrpAsnProLeuThrIleTrpPheTyrluLysGluSerTyrluArgPhe 221
Db 858 CTGGTAACCTGAGTGGAAACCACTTACCGTGTGGTGTCTACCGGACACCTATATCCGCTTT 917
QY 222 SerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsnAla 241
Db 918 TCCAGCGAGCCCTTCTCCCTGAAGAACCTGCACTCAGTGCACCTGTGTGCAACAACCTCC 977
QY 242 ValGlnLysTyrluLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMet 261
Db 978 ATCCGAAGACCTCTGGAGAACTCATGCCATCGGGATCCACTGCTTCGCCGACACACATG 1037
QY 262 TrpThrSerThrArgPheGlnGluTyrluGlnArgGlnGlyArgGlyAlaValTrpGly 281
Db 1038 TGGTCTAGCAGAGTTCCAGGCCCACTCGAGGAGATGGGTGGCCCAAAATGTGTGGTCC 1097
QY 282 SerValIleTyrluProSerMetLysLysAlaIleAlaHisLalaMetLysValAlaGlnAsp 301
Db 1098 ACCATCATCGTCCGTGGCATGAAGGATGCTGTGTATCCACGCACTTCAGACCTCCAGGAC 1157

Qy	302	HisValGluProArgLysAsnSerPheGluLeuTrpGlyAlaAlaPheValLeuGlyArg	321
Db	1158	ACCGTGCATGTCGGAAGCCAGCTTTGAGCTCTATGCGCTGACTTCGTGTTCCGGGAG	1217
Qy	322	AspPheArgProTrpLeuIleGluLeuAsnSerSerProThrMetHisProSerThrPro	341
Db	1218	GACTTCAGCCCTGGCTGATTGAGATCAACGCCAGCCCCCAGATGGCACCCTCCACAGCA	1277
Qy	342	ValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaVal	359
Db	1278	GTCACTGCCCGCTCTGTGCTGGCGTGCAAGCTGACACCCCTGCCGCGTGCATTGACCGG	1337
Qy	360	-----AspArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValVal	377
Db	1338	ATGCTGGACCCCAACTGTGACACAGGAGCCTTTGAGCTCATATATAGCAGCCTGTGTGTG	1397
Qy	378	GluProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgArg	396
Db	1398	GAGGTGCTCAATATGTGGGCATCCGGCTCTGGTAGAGGCTTCACCATCAAGAAGCCC	1457
Qy	397	-----AlaArgArgGlnValLeuProValCysAsnLeu	407
Db	1458	ATGGCGATGTGTATCGCGGATGGGGTTCGCCCCAGCAGTCCTCTGTGACCCAGCGA	1517
Qy	408	-----LysAlaSer-AlaSerLeuLeuAspAlaGlnProLeu---LysAla	421
Db	1518	GGCTCTGGGGAAGCAAGGACTCGGGATCCCTACCCACAGGTCAGCTCTAGGAAGGC	1577
Qy	421	aArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln	440
Db	1578	ACTGGGCGCAGGAGCTTGGGGCACAGTGAGAGCCAGTCTCCATGTCGACCACTTCAGCC	1637
Qy	441	ArgAspLeuGlyLeuLysGluGluLysGlyLeuProLeuAlaLeuAlaProLeuArg	460
Db	1638	CCCGAAGGGGAAGAAAGGCAGGCCGAAAGGGCCACAGCCCTGGTCTGCCCAATCTC	1697
Qy	460	gGlyAlaAla-----GluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAla	477
Db	1698	TGGGAGTGGGATGCCCCCAGCACAGGATGGGCTGCATTTTC-ACCATGACCTTTCTAG	1756
Qy	477	agLYlYValGlu-----LeuProAlaCysProCysArgHisValAspSerGlnAla	494
Db	1757	TGGGGACAGGCACCCACCACCTTGAACAGATTGCCACTGAGTCCGAAGAACCCCGAGC	1816
Qy	494	aProAsnThrGlyValProValAlaGln-----ProAlaLys	506
Db	1817	CCTGGGTAAAGACCATTCCTCCCAAAACCCGAGTGTTCNAGCGATTTATTCCTGCTCT	1876
Qy	506	sSerTrpAspProAsnGlnLeuAsnAlaHisPro-----	517
Db	1877	CCAGGCCCTTCCCAACCACCTGGATCAGCCACCCCAAGAGCCACCAGTAGCAAGTA	1936
Qy	517	-----	517
Db	1937	AAAGCCACTACTACAAAGTATTGTTTAAATAATACAGCCAAATTAGTGGGCACGGTG	1996
Qy	518	-LeuGluProValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGln	537
Db	1997	GTGTGAGCCTGTGGTCCAGGCTACTCGCGAGGCTAATGAGGAT---CGTTGAGCCCGG	2053
Qy	537	gLYlYsGly	540
Db	2054	AGGTCAAGGC	2063

RESULT 9
 AAH16735
 ID AAH16735 standard; cDNA; 2326 BP.
 XX
 AC AAH16735;
 XX
 DT 26-JUN-2001 (first entry)
 XX Human cDNA sequence SEQ ID NO:15920.

CC The invention relates to a new isolated polynucleotide (a human kidney
 CC tumour specific cDNA) comprising any one of the 1855 sequences identified
 CC in the specification (or their complements, degenerate variants,
 CC sequences consisting of at least 20 contiguous residues them, sequences
 CC that hybridise to them under highly stringent conditions or sequences
 CC having at least 75 or 90% sequence identity to the 1855 sequences. Also
 CC included are detecting/determining the presence of cancer in a patient,
 CC stimulating an immune response in a patient; treating kidney cancer in a
 CC patient; an isolated polypeptide encoded by one of the 1855 sequences, an
 CC expression vector comprising the polynucleotide operably linked to an
 CC expression control sequence, a host cell transformed/transfected with the
 CC vector, an isolated antibody (or its antigen-binding fragment) that
 CC specifically binds to the protein, a fusion protein comprising at least
 CC one the proteins, stimulating and/or expanding T-cells specific for a
 CC tumour protein, an isolated T-cell population comprising the T-cells, a
 CC composition comprising a first component (such as a carrier or
 CC immunostimulant) and a second component (comprising one of the
 CC polynucleotides, the polypeptides, an antibody, T-cell or an antigen-
 CC presenting cell that expresses the polynucleotide) and a diagnostic kit
 CC comprising at least one of the oligonucleotides, or at least one antibody
 CC and a detection reagent comprising a reporter group. The polynucleotides,
 CC polypeptides, antibodies and antigen-presenting cells are useful for
 CC detecting the presence of, and treating cancer, particularly kidney
 CC cancer by stimulating and/or expanding T-cells specific for a tumour
 CC protein, and stimulating immune response in a patient. The present
 CC sequence is one of the human kidney tumour specific cDNAs. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from USPTO
 CC at seqdata.uspto.gov/sequence.html?DocID=20030109434.

XX
 SQ Sequence 2538 BP; 555 A; 717 C; 734 G; 532 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 5,59e-57 Length: 2538
 Score: 1020.00 Matches: 220
 Percent Similarity: 55.5% Conservative: 69
 Best Local Similarity: 42.2% Mismatches: 151
 Query Match: 35.7% Indels: 81
 DB: 7 Gaps: 12

US-10-635-977-2 (1-541) x ADS73168 (1-2538)

Qy 54 GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGlu 73
 Db 835 GAGTTTGGTGAAGAGCTCTGTGTCGTCGAGAGTACCTTAGCAACTTGGCCACATG 894
 Qy 74 AspileAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu 93
 Db 895 GACATCGACAGGACTGGAGGCCCGCTGTACCTACCCCGAGGGCTGTGTCCTCTTC 954
 Qy 94 ThrGlnGlnTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
 Db 955 CTCACGCGTACTACCAAGTGTCTCCACGAGGGGAGAACTCAGGCACCTCGACACTCAG 1014
 Qy 114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
 Db 1015 GTCCAGCGCTGTGAGGACATCTCGACGAGCTGCAGGCGGTGTATCCCCAGATACATG 1074
 Qy 134 AspGlyLeuArgAsnIleTrpIleLeuValProAlaAlaLysSerArgGlyArgAspIle 153
 Db 1075 GAAGGGATGCGCAACATCTGGATGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1134
 Qy 154 ValCysMetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLeuSer 173
 Db 1135 ATGTGATGGACCACTCGAGGAGATGTGAAGCTGTGTAACGGCAACCCCGTGTGATG 1194
 Qy 174 ArgAspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThr 193
 Db 1195 AAGCAGCGCAAGTGGTGGTGCAGAAATATATTGAGCGGCCCTCTCATCTTTGGCACC 1254
 Qy 194 LysPheAspileArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
 Db 1255 AAGTTTGACCTCAGACAGTGGTTCCTGGTAATGACTGGAACCCACTTACCGTGTGGTTC 1314

Qy 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
 Db 1315 TACCGCGACAGCTATATCGCTTTTCCACGAGCCCTTCTCCCTGAAGAACCTGGACAAC 1374
 Qy 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
 Db 1375 TCAGTGCACCTGTGCAACCACTCCATCCAGAGCACCTGGAGAACTCATGCCATGGCAT 1434
 Qy 254 ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArg 273
 Db 1435 CCACCTGCTTCCCGCAGACACATGTGCTACCCAGAGGTTCCAGGCCACCTGCAAGGAG 1494
 Qy 274 GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla 293
 Db 1495 ATGGGTGGCCCCAATAGTGTGGTCCACCATCATCGTGGCTGGCATGAAGATGCTGTGATC 1554
 Qy 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
 Db 1555 CACGCACTTCAGACCTCCAGGACACCGTGCAGCTGTCCGAGGCCACAGCTTTGAGCTTAT 1614
 Qy 314 GlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluLeuAsnSerSer 333
 Db 1615 GCGCTGACTTCTGTTTGGGAGGACTTCCAGCCCTGGCTGATTGAGATCAACGCCAGC 1674
 Qy 334 ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353
 Db 1675 CCCACGATGGCACCTCCACAGCAGTCACTGCTCCGCTCTGTGTGGCTGCAAGCTGAC 1734
 Qy 354 ThrIleLysValAlaVal-----AspArgSerCysAspileGlyAsnPheGlu 369
 Db 1735 ACCCTGCGCGTGTTCATTGACCGGATGCTGGACCGCAACTGTGACACAGAGCTTTTGAG 1794
 Qy 370 LeuLeuTrpArgGlnProVal---ValGluPro-ProProPheSerGlySerAspLeuCy 388
 Db 1795 CTCATCTATAGCAGCCCGTTCACCACTTCCCGAGGCTCCACACCAAGGC----- 1843
 Qy 388 sValaGlyValSerValArgAlaArgGlnValLeuProValCysAsnLeuLys 408
 Db 1843 ----- 1843
 Qy 408 sAlaSerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArgGlyProSerAlaMetPr 428
 Db 1844 -----CCAGCTCTCTCCCTTCTCCCTCCGACACAGGCGCCAGCTCTCAGACACAGC 1899
 Qy 428 oAspProAlaGlnGlyProProSerProAlaLeuGlnArgAspLeuGlyLeuLysGluCl 448
 Db 1900 ACAGCAAGCTGTGGGCACTAAGGCCCTGTGACCCACAGGCCAA-----GGCCTTGAAGACTC 1956
 Qy 448 uLys-----GlyLeuProLeuAlaLeuAlaPro---LeuArgGlyAla----- 462
 Db 1957 TACCCAGCGTAAGGTCTTTCATTTCCCTCCACCGAACCTTGATTTCAAGGTGGCACCCA 2016
 Qy 463 -----AlaGluSerGlyAlaAlaGlnProThrArgThr----- 474
 Db 2017 GCATCTGAGACCAAGAGGTGGGCTCGACCTGTGACTCACACCCAGTGACAGTGTCT 2076
 Qy 475 -----LysAlaAlaGlyValGluLeuProAlaCysProCysAr 488
 Db 2077 GAGCAGCGGGTTCAGGGTGGAGGGCAGCAGGAGGGGAGGCTCCAGGCTGG----- 2128
 Qy 488 gHisValAspSerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTr 508
 Db 2129 -----CTGGCACCCCAAGGAGAGGCTG 2151
 Qy 508 pAspProAsnGlnLeuAsn-----AlaHisProLeuGluProVal 522
 Db 2152 GTCTCCCTCAGAGCCCTTCTCCACAGACTTCTGTGATCATCTCCTCTTCTCCCTCT 2211
 Qy 522 uArgGlyLeuLysThrAlaGluGlyAlaLeuArgPro-----ProProGlyGlu 538
 Db 2212 T-----TCACACAGGAGCTCTCTCTCTCTGCTGCTCGAGGCCCGCCAGCTGG 2259

QY 538 Y 538
 Db 2260 A 2260

RESULT 11
 ADW42022
 ID ADW42022 standard; cDNA; 2538 BP.
 AC
 AC ADW42022;
 XX
 XX 24-MAR-2005 (first entry)
 XX
 XX cDNA elevated in kidney tumor cells SEQ ID NO 1765.
 DE
 XX gene therapy; diagnosis; pharmaceutical; cancer; cytostatic;
 KW genitourinary disease; neoplasm; renal tumor; gene; ss.
 XX
 XX Unidentified.
 OS
 XX WO200274237-A2.
 PN
 XX
 XX 26-SEP-2002.
 XX
 XX 19-MAR-2002; 2002WO-US010055.
 PF
 XX
 XX 19-MAR-2001; 2001US-0277245P.
 PR
 XX 21-DEC-2001; 2001US-0343340P.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX
 XX
 PF Algate PA, Mannion J, Gaiger A, Gordon B, Harlocker SL;
 FI
 XX WPI; 2002-759855/82.
 DR
 XX
 XX
 XX New isolated polynucleotides and polypeptides, useful for detecting the
 PT presence of, and treating cancer, particularly kidney cancer by
 PT stimulating T-cells specific for a tumor protein, and stimulating immune
 PT response in a patient.
 PT
 XX
 PS Claim 13; SEQ ID NO 1765; 252pp; English.
 XX
 XX The invention describes a new isolated polynucleotide (I) comprising: any
 CC one of the 1855 sequences identified in the specification; complements or
 CC degenerate variants of (a); sequences consisting of at least 20
 CC contiguous residues of (a); sequences that hybridize to (a) under highly
 CC stringent conditions; or sequences having at least 75 or 90% sequence
 CC identity to (a). Also described are: detecting (M1) or determining the
 CC presence of cancer in a patient; stimulating (M2) an immune response in a
 CC patient; treating (M3) kidney cancer in a patient; an isolated
 CC polypeptide (II) encoded by (I) and comprising, or having at least 70 or
 CC 90% sequence identity to, any one of the 8 sequences identified in the
 CC specification; an expression vector (III) comprising (I) operably linked
 CC to an expression control sequence; a host cell (IV) transformed or
 CC transfected with (III); an isolated antibody (V) or its antigen-binding
 CC fragment that specifically binds to (II); a fusion protein (VI)
 CC comprising at least one (II); an oligonucleotide (VII) that hybridizes to
 CC the nucleotide sequences cited above under highly stringent conditions;
 CC stimulating (M4) and/or expanding T-cells specific for a tumor protein;
 CC an isolated T-cell population (VIII) comprising the T-cells in (M4); a
 CC composition (IX) comprising a first component such as a carrier or
 CC immunostimulant and a second component comprising (I), the polypeptide
 CC encoded by (I), an antibody or its antigen-binding fragment that
 CC specifically binds to (II), (VI), or an antigen-presenting cell that
 CC expresses the polynucleotide; and a diagnostic kit (X) comprising at
 CC least one of the oligonucleotide, or at least one antibody and a
 CC detection reagent comprising a reporter group. The polynucleotides,
 CC polypeptides, antibodies and antigen-presenting cells are useful for
 CC detecting the presence of, and treating cancer, particularly kidney
 CC cancer by stimulating and/or expanding T-cells specific for a tumor
 CC protein, and stimulating immune response in a patient. This sequence
 CC represents a kidney tumour cDNA, expression of which is increased in
 CC kidney tumors.

SQ Sequence 2538 BP; 555 A; 717 C; 734 G; 532 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 5.59e-57 Length: 2538
 Score: 1020.00 Matches: 220
 Percent Similarity: 55.5% Conservative: 69
 Best Local Similarity: 42.2% Mismatches: 151
 Query Match: 35.7% Indels: 81
 DB: 12

US-10-635-977-2 (1-541) x ADW42022 (1-2538)

QY 54 GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGlu 73
 Db 835 GAGTTTGTGATGAAGCTCTGTGCGTGGAGGAGTACCTTAGCAACTTGGCCACATG 894
 QY 74 AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu 93
 Db 895 GACATCGACAAGGACCTGGAGGCCCGCTGTACCTCACCCCGAGGGCTGGTCCCTCTTC 954
 QY 94 ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
 Db 955 CTCACGCGCTACTACCAAGTGTGTCCAGAAAGGGGCGAGAACTCAGGCACCTCGACACTCAG 1014
 QY 114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
 Db 1015 GTCCAGCGCTGTGAGGACATCTGCAGCAGCTGCAGGCCGTGGTACCCCGAGATAGACATG 1074
 QY 134 AspGlyLeuArgAsnIleTrpIleLeuLysProAlaAlaLysSerArgGlyArgAspIle 153
 Db 1075 GAAGGGGATCGCAACATCTGGATCGTGAAGCCAGGAGCAAGTCCCGCGACCGAGGCATC 1134
 QY 154 ValCysMetAspArgValGluGluLeuLeuGluAlaAlaAspHisProLeuSer 173
 Db 1135 ATGTGTCATGGACACCTGGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGGTGATG 1194
 QY 174 ArgAspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThr 193
 Db 1195 AAGGACGGCAAGTGGTGGTGCAGAGTATATTGACGGGCCCTCTCTCATCTTTGGCACC 1254
 QY 194 LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
 Db 1255 AAGTTTGACTCAGACAGTGGTTCTCTGGTAACTGACTGGAACCCCACTTACCGTGTGGTTC 1314
 QY 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
 Db 1315 TACCGCGACAGCTATATCCGCTTTTCACGACGCCCTTCTCCCTGAAGAACCTGGGACAC 1374
 QY 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
 Db 1375 TCAGTGACCTGTGCAACAACTCCATCCAGAGCACTCGAGAACTCATGCCATCGGCAT 1434
 QY 254 ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArg 273
 Db 1435 CCATGTTCCCGCAGACACATGTGCTCTAGCCAGAGGTTCCAGGCCCACTTCGTCAGGAG 1494
 QY 274 GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla 293
 Db 1495 ATGGTGCCGCCAAATGCTTGGTCCACCATCATCGTCCGTCGTCGTCGTCGTCGTCGTCG 1554
 QY 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
 Db 1555 CAGCAGCTTCAGACCTCCAGAGCACCGTGCAGTGTGCGAGGCCAGCTTTGAGTCTAT 1614
 QY 314 GlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluIleAsnSerSer 333
 Db 1615 GCGCTGACTTGTGTTTGGGGAGGACTTCCAGCCCTGGCTGATTGAGATCAACGCCAGC 1674
 QY 334 ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353
 Db 1675 CCCACGATGCAACCTCCACAGCAGTCACTGCCCGGCTGTGTGTGCGTGGCAAGCTGAC 1734
 QY 354 ThrIleLysValAlaVal-----AspArgSerCysAspIleGlyAsnPheGlu 369


```
Db 1735 ACCCTGCGGTGCTCAATGACCGGATGCTGGACCGCACTGTGACACAGGACCTTTGAG 1794
Qy 370 LeuLeuTrpArgGlnProVal---ValGluPro-ProPheSerGlySerAspLeuCy 388
Db 1795 CTCATCTATAAGCAGCCGCTCACCACTCTCCCGAGCTCCACCAAGGC-----1843
Qy 388 sValAlaGlyValSerValArgAlaArgGlnValLeuProValCysAsnLeuLy 408
Db 1843 -----1843
Qy 408 sAlaSerAlaSerLeuLeuAspAlaGlnProLeuLyAlaArgGlyProSerAlaMetPr 428
Db 1844 ----CCAGTCTCTTCCCATGTACTCCGACACAGGCGCAGGTCTCTCAGACGACAGC 1899
Qy 428 oAspProAlaGlnGlnProSerProAlaLeuGlnArgAspLeuGlyLeuLyGluG1 448
Db 1900 ACAGCAAGCTGTGGGCACTAAGGCCCTGTGCACACAGGCAA---GGCCTTGAGGACTC 1956
Qy 448 uLys-----GlyLeuProLeuAlaLeuAlaPro---LeuArgGlyAla-----462
Db 1957 TACCACGCGCTAAGGCTCTCAFTTCCCTCCCAACGACCTTGATTCAAGGTGGCACCCA 2016
Qy 463 -----AlaGluSerGlyValAlaAlaGlnProThrArgThr-----474
Db 2017 GCATCCTGAGCCAAAGAAAGGTGGGCTCGACTGTGACTCACCCAGTGGACAGTGCT 2076
Qy 475 -----LysAlaAlaGlyLysValGluLeuProAlaCysProCysAr 488
Db 2077 GAGCAGCGGTTCAGGCTGGAGGCGACAGGAGGCGAGCTCCAGGCTGG-----2128
Qy 488 gHisValAspSerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTr 508
Db 2129 -----CTGGCACCACCAAGGGAAGAGCTG 2151
Qy 508 pAspProAsnGlnLeuAsn-----AlaHisProLeuGluProValLe 522
Db 2152 GTCTCCCTCAGAAGCCCTTCTCTCCACAGACTTCTGATCATCTCTCCCTCTCT 2211
Qy 522 uArgGlyLeuLysThrAlaGluGlyValAlaLeuArgPro-----ProProGlyG1 538
Db 2212 T-----TCACACCGAGGCTCTGCTCTCTGCTCGAGGCGCCCGAGCTGG 2259
Qy 538 Y 538
Db 2260 A 2260

RESULT 12
ACN37881
ID ACN37881 standard; cDNA; 2553 BP.
XX
AC ACN37881;
XX
DT 18-NOV-2004 (first entry)
XX
DE Tumour-associated antigenic target (TAT) cDNA DNa324273, SEQ ID NO:1057.
XX
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
FN WO2004030615-A2.
XX
PD 15-APR-2004.
XX
PF 29-SEP-2003; 2003WO-US028547.
XX
```

PR 02-OCT-2002; 2002US-0414971P.

(GETH) GENENTECH INC.

Wu TD, Zhang Z, Zhou Y;

WPI; 2004-347921/32.

P-PSDB; ABM80420.

New tumor-associated antigenic target polypeptides and nucleic acids, useful in preparing a medicament for treating or detecting a proliferative disorder, e.g. breast, lung, colorectal, ovarian or prostate cancer or tumor.

Claim 1; SEQ ID NO 1057; 7273pp; English.

The invention relates to human tumour-associated antigenic target (TAT) polypeptides, and their related nucleic acids. The TAT polypeptides are overexpressed in cancer tissues compared to normal tissues, and may thus serve as effective targets for the diagnosis and treatment of cancer in mammals. The invention also relates to nucleic acid and polypeptide sequences at least 80% identical to the TAT nucleic acids and polypeptides; expression vectors and host cells comprising a TAT nucleic acid; an antibody specific for a TAT polypeptide; a peptide or organic molecule which binds to a TAT polypeptide; fusion proteins comprising a TAT polypeptide; and methods and compositions for the treatment or diagnosis of cancer in mammals. TAT polypeptides, nucleic acids, antibodies, antagonists, binding molecules and compositions are useful for diagnosing or treating a cell proliferative disorder associated with increased TAT expression, particularly cancers such as breast cancer, colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder cancer, pancreatic cancer, cervical cancer, cancers of the central nervous system, melanoma and leukaemia. TAT nucleic acids may further be used as hybridisation probes, in chromosome and gene mapping, in chromosome identification and in gene therapy. The present sequence represents a TAT nucleic acid of the invention

SQ Sequence 2553 BP; 575 A; 716 C; 733 G; 529 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5,63e-57	Length:	2553
Score:	1020.00	Matches:	220
Percent Similarity:	55.5%	Conservative:	69
Best Local Similarity:	42.2%	Mismatches:	151
Query Match:	35.7%	Indels:	81
DB:	13	Gaps:	12

US-10-635-977-2 (1-541) x ACN37881 (1-2553)

Qy 54 GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGlu 73

Db 835 GAGTTTGTGGATGAAGCTCTGTGTGCGTGCGAGGAGTACCTTAGCACTTGGCCCATG 894

Qy 74 AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu 93

Db 895 GACATGCACAGGACCTCGAGGCGCCGCTGTACCTCACCCCGAGGCTGGTCTCTCTTC 954

Qy 94 ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113

Db 955 CTCAGCGCTACTACCAAGTGTCTCCAGAGGCGCAACTCAGGCACCTCGACACTCAG 1014

Qy 114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133

Db 1015 GTCCAGCGCTGTGAGGACATCTCGACAGCTGCGAGGCGGTGTATCCCATAGACATG 1074

Qy 134 AspGlyLeuArgAsnIleTrpIleIleLeuPheProAlaLysSerArgGlyArgAspIle 153

Db 1075 GAAGGGATCGCAACATCTGATCGTGAAGCAGGAGCCAAAGTCCCGCGGACGAGGCATC 1134

Qy 154 ValCysMetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLeuSer 173

Db 1135 ATGTGCATGGACCACTGGAGGAGATGCTGAAGCTGTGTGAACGCAACCCCGTGTGTATG 1194

XX OS Homo sapiens.
 XX PN WO200153312-A1.
 XX PD 26-JUL-2001.
 XX PF 26-DEC-2000; 2000WO-US034263.
 XX PR 23-DEC-1999; 99US-00471275.
 XX PR 21-JAN-2000; 2000US-00488725.
 XX PR 25-APR-2000; 2000US-00552317.
 XX PR 20-JUN-2000; 2000US-00598042.
 XX PR 19-JUL-2000; 2000US-00620312.
 XX PR 03-AUG-2000; 2000US-00653450.
 XX PR 14-SEP-2000; 2000US-00662191.
 XX PR 19-OCT-2000; 2000US-00693036.
 XX PR 29-NOV-2000; 2000US-00727344.
 XX (HYSE-) HYSEQ INC.
 XX PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao Q;
 XX PI Zhou P, Goodrich R, Drmanac RT;
 XX XX
 DR WPI: 2001-442253/47.
 DR P-PSDB; AAM39450.
 XX
 XX Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.
 XX
 XX Claim 1; SEQ ID NO 809; 10078pp; English.
 XX
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
 CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemoractic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification
 XX
 XX Sequence 3001 BP; 662 A; 854 C; 846 G; 639 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 6,79e-57 Length: 3001
 Score: 1020.00 Matches: 220
 Percent Similarity: 55.5% Conservatives: 69
 Best Local Similarity: 42.2% Mismatches: 151
 Query Match: 35.7% Indels: 81
 DB: 4 Gaps: 12
 US-10-635-977-2 (1-541) x AAI58606 (1-3001)
 Qy 54 GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlnLeuHisGlu 73
 Db 1304 GAGTTTGTGGTGAAGAGTCTGTGTGCGTGGAGGAGTACCTTAGCAACTTTGGCCCATATG 1363
 Qy 74 AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTTPGluAspLeu 93
 Db 1364 GACATCGACAAAGGACCTGGAGGCCCGCTGTACCTCACCCCGGAGGCTGTCTCTTC 1423
 Qy 94 ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
 Db 1424 CTCAGCGCGCTACTACCAAGTGGTCCACGAGGGGCGAGAACTCAGGCACCTCGACACTCAG 1483

Qy 114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
 Db 1484 GTCCAGCGCTGTGAGGACATCTCTGCAGCAGCTGCGAGCGCTGGTACCCAGATAGACATG 1543
 Qy 134 AspGlyLeuArgAsnIleTrpIleIleLysProAlaLysSerArgGlyArgAspIle 153
 Db 1544 GAAGGGGATCGCAACATCTGGATCTGAAGCCAGGAGCAAGTCCCGCGGAGGAGCATC 1603
 Qy 154 ValCysMetAspArgValGluGluLeuLeuAlaAlaAspHisProLeuSer 173
 Db 1604 ATGTGCATGGACACCTGGAGGAGATGCTGAAGCTGGTGAACGCGCAACCCCGTGGTGTG 1663
 Qy 174 ArgAspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThr 193
 Db 1664 AAGGACGGCAAGTGGTGGTGCAGAGTATATTGAGCGGCCCTCTCTCATCTTTGGCACC 1723
 Qy 194 LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
 Db 1724 AAGTTTGACCTCAGACAGTGGTTCCTGGTAACTGACTGGAACCCCACTTACCGTGTGGTTC 1783
 Qy 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
 Db 1784 TACCGCAGCAGTATATCCGCTTTTCCACGACGCCCTCTCTCCGTAAGAACCTTGGACAAC 1843
 Qy 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
 Db 1844 TCAGTGCACCTGTGCNACAACTCCATCCAGAGCAGCTCGAGNACTCATGCCATCGGCAT 1903
 Qy 254 ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArg 273
 Db 1904 CCACGTCTTCGCGCAGACAAACATGTGTCTAGCCAGAGGTTCACGGGCCCACTTCGAGGAG 1963
 Qy 274 GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla 293
 Db 1964 ATGGTGCCGCCAAATGCTTGGTCCACCATCATGCTGCGTGGCAGGAGATGCTGGATC 2023
 Qy 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
 Db 2024 CACGCACTTCAGACCTCCAGAGCAGCCGTCGAGTGTCCGAGGCCAGCTTTGAGCTCTAT 2083
 Qy 314 GlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluLeuAsnSerSer 333
 Db 2084 GGGCGTGTCTGTGTTCGGGGAGGACTTCCAGCCCTGGCTGATTGAGATCAACGCCAGC 2143
 Qy 334 ProThrMetHisProSerThrProValThrAlaGlnCysAlaGlnValGlnGluAsp 353
 Db 2144 CCCACATGGCACCCCTCCACAGCAGTCACTGCCCGGCTCTGTGCTGGCGTCAAGCTGAC 2203
 Qy 354 ThrIleLysValAlaVal-----AspArgSerCysAspIleGlyAsnPheGlu 369
 Db 2204 ACCCTGCGGTGTGTCATTGACCGGATGCTGGACCGCAACTGTGCACAGGAGCTTTGAG 2263
 Qy 370 LeuLeuTrpArgGlnProVal---ValGluPro-ProProPheSerGlySerAspLeuCy 388
 Db 2264 CTCATCTATAAGCAGCCCGTCCACACTTCCCGAGCTCCACACCAAGGC----- 2312
 Qy 388 sValAlaGlyValSerValArgAlaArgGlnValLeuProValCysAsnLeuLy 408
 Db 2312 ----- 2312
 Qy 408 sAlaSerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArgGlyProSerAlaMetPr 428
 Db 2313 -----CCAGTGCCTTCTCCCATGTACTCCGACACCGGCCAGGCTCTCAGACGACAGC 2368
 Qy 428 oAspProAlaGlnGlyProProSerProAlaLeuGlnArgAspLeuGlyLeuLysGluGl 448
 Db 2369 ACAGCAAGTGGTGGGCACCTAAGGCCCTGTGCGACCAACAGGCAA---GGCCTTGAGGACTC 2425
 Qy 448 uLys-----GlyLeuProLeuAlaLeuLeuAlaPro---LeuArgGlyAla----- 462
 Db 2426 TACCACCGGCTAAGGTCTTCATTTCCCTCCACCGAACCTTGATTTCAAGGTGGCACCCA 2485
 Qy 463 -----AlaGluSerGlyGlyAlaAlaGlnProThrArgThr----- 474

Db 2486 GCATCTGAAGCAAGAAAGGTGGCGCTCGAAGCTGTGACTCAACCCAGTGGACAGTGCT 2545
 QY 475 -----LysAlaAlaGlyLysValGluLeuProAlaCysProCysAr 488
 Db 2546 GAGCAGCGGTGAGGGCTGAGGGACAGCGAGGGAGCTCCAGGCTGG----- 2597
 QY 488 gHisValAepSerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTr 508
 Db 2598 -----CTGGCACCCCAAGGGAAGAGCTG 2620
 QY 508 pAspProAnGlnLeuAsn-----AlaHisProLeuGluProValLe 522
 Db 2621 GTCCTCCTCAGAGGCCCTCTCTCCACAGACTTCTGATCATCTCCCTCTCTCCCTCTCT 2680
 QY 522 uArgGlyLeuLysThrAlaGluGlyAlaLeuArgPro-----ProProGlyG1 538
 Db 2681 T-----TCACACCGAGGCTCTGCTCTCTGTCCTCGAGGCCGCCAGCTGG 2728
 QY 538 Y 538
 Db 2729 A 2729

RESULT 15

ADQ98824

ID ADQ98824 standard; cDNA; 3001 BP.

AC ADQ98824;

DT 23-SEP-2004 (first entry)

DE DNA encoding human GPCR-like protein seqid 494.

KW ophthalmological; immunomodulatory; cytostatic; antiatherosclerotic;
 KW antidiabetic; GPCR-like protein; ophthalmic disorder;
 KW neurological disorder; immunological disorder; nephritic disorder;
 KW hormonal dysfunction; cancer; atherosclerosis; diabetes;
 KW molecular weight marker; food supplement; human; ss.

OS Homo sapiens.

PN US6569662-B1.

PD 27-MAY-2003.

PF 19-JUL-2000; 2000US-00620312.

PR 21-JAN-2000; 2000US-00488725.

PR 25-APR-2000; 2000US-00552317.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Zhou P, Drmanac RT;

XX WPI; 2001-442255/47.

DR New G-protein-coupled receptor-like polypeptides and polynucleotides,
 PT useful for treating diseases of ophthalmic, neurological, immunological
 PT and nephritic systems and hormonal dysfunction, cancer, atherosclerosis
 PT and diabetes.
 XX Example 2; SEQ ID NO 494; 92pp; English.

The invention describes an isolated polynucleotide (I) comprising a fully
 defined (SI) of 749, 3188, 2484, 1169, 2936, 1467, 5773, 5714, 4041,
 CC 1372, 3996, 3945, 1788, 585, 1782, 927, 5714 or 2282 nucleotides as
 CC given in the specification, its translated or protein coding portion, its
 CC extracellular portion or its active domain. The GPCR-like polypeptides
 CC and polynucleotides are useful for the treatment of diseases of
 CC ophthalmic, neurological, immunological and nephritic systems. They may
 CC also be used to treat hormonal dysfunction, cancer, atherosclerosis and
 CC diabetes. The antibodies are useful for detecting or quantitating the
 CC polypeptide in tissue. The polypeptides can also be used as molecular

CC weight markers and as a food supplement. This sequence represents a human
 CC polynucleotide of the invention.

SQ Sequence 3001 BP; 662 A; 854 C; 846 G; 639 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6,79e-57 Length: 3001
 Score: 1020.00 Matches: 220
 Percent Similarity: 55.5% Conservative: 69
 Best Local Similarity: 42.2% Mismatches: 151
 Query Match: 35.7% Indels: 81
 DB: 5 Gaps: 12

US-10-635-977-2 (1-541) x ADQ98824 (1-3001)

QY 54 GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGlu 73
 Db 1304 GAGTTTGGTGAAGCTGTGTGTCGCGAGGAGTACCTTAGCAACTTGGCCCAATG 1363
 QY 74 AspileAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu 93
 Db 1364 GACATCGACAGGACCTGGAGGCCCGCTGTACCTCACCCCGAGGGCTGGTCCCTCTTC 1423
 QY 94 ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
 Db 1424 CTCCAGCGCTACTACCAAGTGTCTCCAGAGGGGAGAACTCAGGCACCTCGACACTCAG 1483
 QY 114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
 Db 1484 GTCCAGCGCTGTGAGGACATCTCTGAGCGAGGCGGTGTTACCCAGATAGACATG 1543
 QY 134 AspGlyLeuArgAsnIleTyrIleLysProAlaAlaLysSerArgGlyArgAspIle 153
 Db 1544 GAAGGGATCGCAACATCTGGATCGTGAAGCAGCAGGACCAAGTCCCGGGAGCGGCATC 1603
 QY 154 ValCysMetAspArgValGluGluLeuLeuGluLeuAlaAlaAspHisProLeuSer 173
 Db 1604 ATGTGCATGGACCACTCGAGGAGATGCTGAAGCTGTGTGAACCGCAACCCCGTGTGATG 1663
 QY 174 ArgAspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuLysCysAspThr 193
 Db 1664 AAGGACGCAAGTGGTGGTGCAGAAATATTTAGCGCGCCCTCTCTATCTTTGGCACC 1723
 QY 194 LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
 Db 1724 AAGTTTGACCTCAGACAGTGTCTCTGTACTGACTGGAACCCACTTACCGTGTGGTTC 1783
 QY 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
 Db 1784 TACCGCAGACAGTATATCCGCTTTTCCAGCAGCGCTTCTCCCTGAAGAACCTGGACAAC 1843
 QY 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
 Db 1844 TCAGTGACCTGTGCAACACTCCATCCAGAAAGCACTGGAGAACTCATGTCATCGGCAT 1903
 QY 254 ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArg 273
 Db 1904 CCATGTCTCCGCCAGACACATGTGTCTAGCAGAGGTTCAGGCCACCTCGAGGAG 1963
 QY 274 GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla 293
 Db 1964 ATGGGTGCCCAATGCTTGTTCACCATCATCTGCTGCTGCATGTAAGGATGCTGTGATC 2023
 QY 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
 Db 2024 CACCACTTTCAGACCTCCAGGACACCGTGCAGTGTGCGAAGGCCAGCGCTTTGAGCTCTAT 2083
 QY 314 GlyAlaAspValLeuLeuGlyArgAspPheArgProTrpLeuIleGluLeuAsnSerSer 333
 Db 2084 GGCCTGACTTCGTGTTCGGGGAGGACTTCCAGCCCTGGCTGATTGAGATCAAGCCACG 2143
 QY 334 ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, **using frame_plus.p2n-model**

Run on: April 4, 2006, 09:36:08 ; Search time 7900 Seconds

(without alignments)

3892.697 Million cell updates/sec

Title: US-10-635-977-2

Perfect score: 2854

Sequence: 1 MASSILKVVSHQSCSRSSR.....LRGLKTABGALRPPPGKGKS 541

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEVEL=xlh
-Q=/abss/ABSSWEB spool/US10635977/runat_04042006_103531_9993/app query.fasta_1
-DB=GenEmbl -QFMT=fastcap -SURFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss03h
-USER=US10635977 @CGN_1_1_4939 @runat_04042006_103531_9993 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_scs:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	1802	63.1	101270	8	HS355C18
2	1455	51.0	817	6	CQ724907
3	1075	37.7	4238	8	HS805098
					AL022327 Human DNA
					CQ724907 Sequence
					AL833939 Homo sapi

4	1070.5	37.5	2380	6	AX301197	Sequence
5	1056.5	37.0	2097	8	BC098298	Homo sapi
6	1054	36.9	2326	6	BD158727	BD158727 Primer fo
7	1054	36.9	2326	6	AX881015	Sequence
8	1054	36.9	2326	8	AK023960	Homo sapi
9	1020	35.7	2553	8	HS800637	Homo sapi
10	1020	35.7	3001	6	AR339003	Sequence
11	1001.5	35.1	1769	8	BC099735	Homo sapi
12	1000.5	35.1	1743	8	BC098361	Homo sapi
13	983.5	34.5	2848	6	AX834642	Sequence
14	983.5	34.5	2848	8	AK097236	Homo sapi
15	960.5	33.7	1832	4	BT021865	Bos tauru
c 16	942	33.0	205949	14	AC134940	Rattus no
17	942	33.0	218249	14	AC097425	Rattus no
18	940.5	33.0	5282	6	CQ842940	Sequence
19	940.5	33.0	5282	8	AK125875	Homo sapi
20	922	32.3	1684	8	AF078842	Homo sapi
21	880	30.8	1897	9	BC006830	Mus muscu
22	856	30.0	163783	9	AC119559	Sequence
c 23	772	27.0	3828	6	CQ850619	Sequence
c 24	772	27.0	3828	8	AK127786	Homo sapi
25	733	25.7	2979	6	CQ595968	Sequence
c 26	556.5	19.5	5728	6	CQ595967	Sequence
27	556.5	19.5	76854	2	AC003052	Sequence
28	556.5	19.5	86398	14	AC017785	Drosophill
c 29	556.5	19.5	167201	2	AC092397	Drosophill
c 30	556.5	19.5	259718	2	AE003614	Drosophill
31	536.5	18.8	2543	6	CQ595656	Sequence
32	535.5	18.8	3914	2	BT021443	Drosophill
33	524	18.4	1813	9	BC097453	Rattus no
34	514.5	18.0	4615	6	CQ595655	Sequence
c 35	514.5	18.0	76854	2	AC003052	Sequence
c 36	514.5	18.0	86398	14	AC017785	Drosophill
37	514.5	18.0	167201	2	AC092397	Drosophill
38	514.5	18.0	25965	9	AC153589	Mus muscu
39	514.5	18.0	259718	2	AE003614	Drosophill
c 40	501.5	17.6	224077	14	AC120010	Drosophill
c 41	486	17.0	754	10	BV579068	Mus muscu
42	480.5	16.8	1958	6	BD160681	Primer fo
43	480.5	16.8	1958	8	AX884044	Sequence
44	480.5	16.8	1958	8	AK024110	Homo sapi
c 45	478.5	16.8	145435	14	AC026685	Homo sapi

ALIGNMENTS

RESULT 1

HS355C18/c

LOCUS

DEFINITION

Accession

Version

Keywords

Source

Organism

Reference

Authors

Title

Journal

Comment

HS355C18 101270 bp DNA linear PRI 18-MAY-2005
Human DNA sequence from clone RP3-355C18 on chromosome 22q13.3
Contains the KIAA0027 gene, ESTs, STSS, GSSs and seven putative CpG
islands, complete sequence.

AL022327 GI:5304851
AL022327.17 HTG; CpG island; KIAA0027.7.
HTG; CpG island; KIAA0027.7.
Homo sapiens (human)

Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 101270)

Cobley, V.
Direct Submission
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vga@sanger.ac.uk

Clone requests: clonerequest@sanger.ac.uk
On Jun 30, 1999 this sequence version replaced gi:5262834.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:

Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/Celegans/wormpep This sequence
was generated from part of bacterial clone contigs of human

chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>
RP3-355C18 is from the library RPCI-3 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pCIPAC2
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: vega@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
Location/Qualifiers
1. .101270
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="RZPD:RPCIP704C18355"
/db_xref="taxon:9606"
/chromosome="22"
/map="q13.3"
/clone="RP3-355C18"
/clone_lib="RPCI-3"

FEATURES

source

misc_feature

misc_feature

misc_feature

gene

mRNA

CDS

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

gene

mRNA

polyA_site

polyA_signal

CDS

misc_feature

misc_feature

ORIGIN

Alignment Scores:
Pred. No.: 1,3e-83 Length: 101270
Score: 1802.00 Matches: 438
Percent Similarity: 32.2% Conservative: 1
Best Local Similarity: 32.1% Mismatches: 3
Query Match: 923
DB: 6
Gaps: 3
US-10-635-977-2 (1-541) x HS355C18 (1-101270)
Qy 100 LeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAla 119
Db 4722 CTTTGTAGTGGCGATGCTTTTCATCTCCATTCAGAAATTTACTTTTCGCAATGCCAGGCT 47163
Qy 120 LeuLeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIle 139
Db 47162 CTCTCTGAATAGATCAGCTGTGTGAACCCCTCAGACGACATTCAGCGGCTCCGGAACATC 47103
Qy 140 TrpIleIleLysProAlaLalaLysSerArgGlyArg----- 151
Db 47102 TGGATTATAAGCCCGCGCCAAAGTCCCGGGGCG-AGTGTAGTCCCTCGTGGCGATGC 47044
Qy 151 ----- 151
Db 47043 TCCCTGCACCGCGTTTCATAGCCACCAACGCCACACTTGATGCCAGGCGGTGAGTGGAG 46984
Qy 151 ----- 151
Db 46983 GGACAGTTTTCAGGCCCAACACACAGGAGGAAGGAGGGAGCAGAACGCCGAGCA 46924
Qy 151 ----- 151
Db 46923 GGCGTGGGGGTTCAGGACTCAGCGCTGGACAGAGCTGGAGGGCGTCTTTCACAGCATGTGT 46864
Qy 151 ----- 151

52740
/note="Tandem repeat. Assembly not confirmed by digest"
52620
/note="Tandem repeat. Assembly not confirmed by digest"
70150
/note="Tandem repeat. Assembly not confirmed by digest"
complement(72100..98325)
/locus_tag="RP3-355C18.1-001"
complement(join(72100..74366,76743..76907,81142..81264,83236..83292,86925..87041,89550..89621,90110..90211,92627..92728,93053..93106,95793..95882,97435..97670,98064..98325))
/locus_tag="RP3-355C18.1-001"
complement(72106)
/locus_tag="RP3-355C18.1-001"
complement(72118..72123)
/locus_tag="RP3-355C18.1-001"
complement(join(74292..74366,76743..76907,81142..81264,83236..83292,86925..87041,89550..89621,90110..90211,92627..92728,93053..93106,95793..95882,97435..97611))
/locus_tag="RP3-355C18.1-001"
/standard_name="ORTHUMP00000028512"
/codon_start=1
/protein_id="CAI42687.1"
/db_xref="GI:57208208"
/translation="MTQEPFREELAYDRMPTLERGRODPASVAPDAKPSDLQSLKRLP
PCFSHTKWTWFSVLMSGSLLVTSGLTSGFSYILGNVFPABMDYLRCAGSGCIPSAIVSFTVSR
RNVNVPNFIQILFVSTFAVTTTCLWFGCKVLNPSAININFNLLILLLLELLMAATV
IIAARSEEDCKKKGSDSANILDEVPFPAVLKSYVEVIAIGSAVLGGIILAN
VDDSGPHLSVTFWILVACFPAIAASHAAECPSKCLVEVLIAISLTSPLLTAS
GYLSFSIMRIVEMFKDYPPIKPSVDLLELLVLLQAGLNTGTAIQCVRPKVSAR
LQASWDVTQNGPQERLAGEVARSPLEKPKERAWRVVQMAQ"
80430
/note="Tandem repeat. Assembly not confirmed by digest"
101270
/note="Clone_right_end: RP3-355C18"

Db	46863	TCCGATCAGAGGCAGGATCCGGGTTGATGGGTTAATCAGGATTTCTCTCTTTGTAATCA	46804
Qy	151	-----	151
Db	46803	CAAACTGGTCATGCTTCTCTTTTAGTTTATTTATTTATTTATTTATTTTAGTTTTT	46744
Qy	151	-----	151
Db	46743	TGAACGGAGTTTCACTCTGTGTTGCCAGGCTGAAGTGAGTGCGCGCAATCTTGCTCAC	46684
Qy	151	-----	151
Db	46683	TGCAACCTCCACCTTCCAGGTTTAAGCAATTTCTCTGCTTACCTTCCCAAGTAGTGGG	46624
Qy	151	-----	151
Db	46623	ACTACAGGCGCCACACACCGGCTAATTTTGTGATTTTATTTAGTAGATGGAGTTTCA	46564
Qy	151	-----	151
Db	46563	CCATGTTGGCAGGCTGCTTTGAACCTCCGACCTCAAGTGATTCACCTGCTCGGCTC	46504
Qy	151	-----	151
Db	46503	CCAAAGTGTGGGATTACAGGCATAAGCACCAATGCCATCTCTGGCTCCCTTTTAA	46444
Qy	151	-----	151
Db	46443	AGTGTCCTGTCCCTGCCCTTCACTGATCTCTGTGTCCTCTGGGTCTCAGAGGCTG	46384
Qy	151	-----	151
Db	46383	GTTCGTTGGGGAGAGCCTGGGGACTCAGGCACTGGCTTTAGCCTCATCATGAGCG	46324
Qy	151	-----	151
Db	46323	TAAGCTTCCCTGCTCCCATGCGCCCAACCAACCATGCTGCTCTGGCTTGGACAA	46264
Qy	151	-----	151
Db	46263	AGCCAGTGCAGTCTCAGGGCCAAAGCTGGACCTGCGCCCTGCGAGGCTCAGGCTC	46204
Qy	152	-----AspIleValCys 155	
Db	46203	CGCTGCACAGTGGCTGTGCATGCTGGGTTGAGCGTGGCTCTTGCAGACATAGTGC	46144
Qy	156	MetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAsp	175
Db	46143	ATGGACCGTGTGGAGGATCCTGGAGCTGGCAGCTGGACACACCTCTTTCCAGGGAC	46084
Qy	176	AsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPhe	195
Db	46083	AACAAGTGGTGTCCAGAGTACATCGAGACCGCTGCTCATCTGTGACACCAAGTTC	46024
Qy	196	AspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLys	215
Db	46023	GACATCAGACAGTGGTCTCTCGTCACGGACTGGAACCCCTGACCATCTGGTTCTACAG	45964
Qy	216	GluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAsp	232
Db	45963	GAGAGTTACTTGGGTTCTCACTCAGGCTTCTCCCTGGACAAGCTGGACAGTGCAGTG	45904
Qy	232	-----	232
Db	45903	CGTGGGCGACCTGGACACCTGCGGCCAGGGGAAATGGGCTGCTGCTGCATGCTGGAGCA	45844
Qy	232	-----	232
Db	45843	GCCGAGTGAGAGCAGCAGCGGCGAGGGCCCTCACTGCCACCTGCCACACGAGAGGC	45784
Qy	232	-----	232
Db	45783	ACAGATGGGGAACCTGAGGCTTGGGAGGCTGATGATCTCCCCGAGGGACAGGCCAGGCC	45724
Qy	232	-----	232
Db	45723	CACCGCCTTCAAAGCCAGCGCCTACACCCACCGCTCTGTGATCATCTCTGGAATAATC	45664
Qy	232	-----	232
Db	45663	TTAAAGGCACCATTTTGCTGTGCCAGGGGATTTGAATGTTGTGGGAGCACCCTGGG	45604
Qy	232	-----	232
Db	45603	CCCTGGCGCAGGGGAGCGGCTGTGGGCTGCTGGTGAGGCTCAGATATCCGCAAGTG	45544
Qy	232	-----	232
Db	45543	TGCTGCCAGGCTGGATGGACCCACCTCACCCGGGTCCCGCTGGGAGGAGGCGCA	45484
Qy	232	-----	232
Db	45483	GAGCCTGTGGAGTCTCTCAGGAGTGTCTGTGCACACAGCGAGCGCTGCCACCTCTGTCT	45424
Qy	232	-----	232
Db	45423	GCACCAAGCTTCTCTGACTATTGTGTGGCCCTTGTGTTGGTGTGAGGCTAGGGGACC	45364
Qy	232	-----	232
Db	45363	TGACGGCTGGGAGGAGGGGCTGTGCTTTTACGAGTGAGAATCCAGGGAGCCTGAGA	45304
Qy	232	-----	232
Db	45303	GTCACGCGACACAGACCTTGACGGGAGTTAACTCGGGCCAGGGAGCGCTGGCGGTCA	45244
Qy	232	-----	232
Db	45243	TGCTGTGGTGGTCAATGCTCGAAGTCACACCCAGAGAGCGCTGGGAGGAGTGGCCAGG	45184
Qy	232	-----	232
Db	45183	GCTACGACCTCACCTGTCTACTGTCTGTGTCAAAGTGGGCTGGAGGGGTCCAGGAAC	45124
Qy	232	-----	232
Db	45123	TCACCTGTGCAAAAGCAGCTCCAGGCGAGTCAGAGCAGTGAGAAGCAGCGCCCTGCCAGCG	45064
Qy	232	-----	232
Db	45063	CCCATTTCTGGGCTTCCAGCTGTGAGCTGTGCGCTGGCCATGCCATGGCCTCTCTGAG	45004
Qy	232	-----	232
Db	45003	CCCTGTTTCTCATCTGGGACAGTAACGCTCATCCAGCTGGGCGCTCCAGCAGAGGT	44944
Qy	232	-----	232
Db	44943	TGGAGTCTCTCAGGCGAGGTTCTCTAGACGAGCTGGTGGGCTCAGCACGGCTGGGAG	44884
Qy	232	-----	232
Db	44883	GACCGGCTTGGGTGCCCTCTCTGGGCCAAGGAGGCTGAGCACCTCGTCCCCC	44824
Qy	233	---SerAlaIleHisLeuCysAsnAlaValGlnLysTyrLeuLysAsnAspValGly	251
Db	44823	TGCAGCGCATCCACTGTGCAACACCGCTCCAGAGTACCTGAAGATGATGTGGGC	44764
Qy	252	ArgSerProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeu	271
Db	44763	CGCAGCCCCCTGCTGCCCGCACACAAACATGTGGACAGCACCAAGTTCCAGGAGTACCTG	44704
Qy	272	GlnArgGlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysAla	291
Db	44703	CAGCGCCAGGGCGCTGGCGCTGTGGGCGAGCGTCACTACCCGCTCCATGAAGAAGGCC	44644

```
QY 292 IleAlaHisAlaMetIysValAlaGlnAspHisValGluProArgIysAsnSerPheGlu 311
Db 44643 ATGCGCCAGCCCATGAAGGTGGCCAGGACCAAGTGGAGCCTCGCAAGAACAGCTTTGAG 44584
QY 312 LeuTyrGlyAlaAspPheValLeuGlyArgAspPheArgProTyrLeuIleAsn 331
Db 44583 CTCTACGGGGCTGACTTCGTCCTTTGGAGGAGACTTCAGGCCCTGGCTGATCGAGATCAAT 44524
QY 332 SerSerProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGln 351
Db 44523 TCCAGCCCCACCATGCACCGCTCCACCGGTCACGGCCAGCTGTGTGCACAGGTGCAG 44464
QY 352 GluAspThrIleIysValAlaValAlaAspArgSerCysAspIleGlyAsnPheGluLeu 371
Db 44463 GAGGACACCATCAAGGTGGCCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCCTG 44404
QY 372 TTPArg----- 373
Db 44403 TGGAGCAGGTGAGCCACGCGCCCGCTGGGGACTTTGGTGGCGCGCCGACATGTGGATG 44344
QY 373 ----- 373
Db 44343 CTTTGGCCACTGCTCACTGGCGAGGTGGAGCTGCAGGCGCGCAGTGCAGCGGGTCTA 44284
QY 373 ----- 373
Db 44283 TGGAGGACACCGATCCTGCTCTGTGACCATGCTCGGTCCGCTGAGGGGCCAGGCGCTGCC 44224
QY 373 ----- 373
Db 44223 CTGCTCTCCATCGGGCTCAGTCCGCTGCTCGGCACCGAGGAGGGGTACTCCAGGGGTG 44164
QY 373 ----- 373
Db 44163 CGGGCGAGGGCCCGGCTCACTCAGGCTCAGTGCCTCTGTCCAGCTTTCTCCCTCTTG 44104
QY 373 ----- 373
Db 44103 TTCCAGACCTCACCCAGTCTCTGACTGTGTGTAGAGTGGGGGTCCATGCCCGCAGCCTC 44044
QY 373 ----- 373
Db 44043 CAGGCCACCTATGCTGCTCCTTTGCTTGGCTTGCCTGAGTGCTCCTCCACTCTCAGGATCT 43984
QY 373 ----- 373
Db 43983 TTGCGGGGCTCCCTGTGCTGCGGATGCTGTGAGTCAATTGGAGCGGTGGCCAAATGCC 43924
QY 373 ----- 373
Db 43923 TCTCTCTTGGGGCTCTCTGTGGCTAACTCGGGAAGAGACGCTGAGTCCCTGGGTGT 43864
QY 373 ----- 373
Db 43863 GGGGCCCCCAGGCTGACCCCGAAGCCTCAGGAAGCCCTGTGTTCCCGCAGGTTCTCCAG 43804
QY 373 ----- 373
Db 43803 AGCTGGGGGCTCTCTGGGGTCTGCTGGGACCCAGGTGTTTCACTGTTCACTCTCTGAG 43744
QY 373 ----- 373
Db 43743 TCAGTGGGGTCTGTGCCCCGACCAAGCCCTCTCTGCACAGACCCCTCTCTGCGAGGCC 43684
QY 374 -----GlnProValValG1 378
Db 43683 CCCCTATCCCGGCCACAGCCAGCACCCCAACCTGTTCTTCCCGACAGCGGTGGTGA 43624
QY 378 uProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaAr 398
Db 43623 GCCGCCCCCATTCAGCGGGTCCGACCTCTGCTGGCGGGGTCAAGTGTGAGGAGGCGAG 43564
```

```
QY 398 gArgGlnValLeuProValCysAsnLeuIysAlaSerAlaSerLeuLeuAspAlaGlnPr 418
Db 43563 GAGGCAGGTGCTGCGCGTCTGCAACTCAAGGCTCGGCTCGCTGTTGGACGCGCAGCC 43504
QY 418 oLeuIysAlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAl 438
Db 43503 GCTGAAGGCACGGGGCCCTCGGCCATGCGACACCTGCGGAGGAGCCGCCCATCACCAGC 43444
QY 438 alLeuGlnArgAspLeuGlyLeuIysGluIysGlyLeuProLeuAlaLeuLeuAlaPr 458
Db 43443 TCTCCAGCGGACTTGGGACTGAAGGAAGAGAGGGGCTCCCTCGCTTGTCTGGCACC 43384
QY 458 oLeuArgGlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrIysAlaAlaG1 478
Db 43383 CTTTAGGGGGGCGAGCGAGCGGTGAGCGCGCACAGCCACCCGACCAAGCTGCTGG 43324
QY 478 yLysValGluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrG1 498
Db 43323 GAAGGTGGAGCTCCCGGCTGCGCCCTGTGCGCCACGTGAGTCAAGTCAAGCCCAACACCGG 43264
QY 498 yValProValAlaGlnProAlaIysSerTrpAspProAsnGlnLeuAsnAlaHisProLe 518
Db 43263 TGTCCCGGTAGCCAGCCCGCAAAAGCTGGGATCCAAACCACTAAATCGGCACCCGCT 43204
QY 518 uGluProValLeuArgGlyLeuIysThrAlaGluGlyAlaLeuArgProProGlyG1 538
Db 43203 GGAGCCTGTGCTGCGGGGCTGNAGACAGAGGGGCGGCTGCGTCCGCGCGCGGAGG 43144
QY 538 yLysGlySer 541
Db 43143 AAAAGGTAAC 43134

RESULT 2
LOCUS CQ724907
DEFINITION Sequence 10841 from Patent WO02068579.
ACCESSION CQ724907
VERSION CQ724907.1 GI:42285764
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of humanexons or transcripts, for detecting expression and other uses thereof
JOURNAL Patent: WO 02068579-A 10841 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source 1..817
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.: 9.97e-68 Length: 817
Score: 1455.00 Matches: 272
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 51.0% Indels: 0
DB: 6 Gaps: 0

US-10-635-977-2 (1-541) x CQ724907 (1-817)

QY 103 GlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeuLeuAsn 122
Db 2 GCGGATGCTTTCATCTCAATTCAAGAAATTTACTTTTCGAGTGCCAGGCTCTGCTGAAT 61
QY 123 ArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIleIle 142
```



```
Db 1687 CAGCAGCTGAGCGCGTGTACCCAGATAGACATGGAAGGGATCGCAACATCTGGATC 1746
QY 142 IleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGlu 161
Db 1747 GTGAAGCCAGGAGCCAAAGTCCCGTGACGAGGACATCATGTGATGACCAACCTGGAGGAG 1806
QY 162 IleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValGln 181
Db 1807 ATGCTGAAGCTGTGTAACGCGCAACCCCGTGGTGTATGAAGACGCGGAGTGGGTGGCAG 1866
QY 182 LysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPhe 201
Db 1867 AAGTATATTAGCGCGCCCTCTCTCATCTTTGGCACCAAGTTTGACCTCAGACAGTGGTTC 1926
QY 202 LeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPhe 221
Db 1927 CTGTAACCTGACTGGAACCCACCTACCGTGTGTTCTACCGCAGCAGCTATATCCGCTTT 1986
QY 222 SerThrGlnArgPheSerLeuLysLysLeuSerAlaIleHisLeuCysAsnAsnAla 241
Db 1987 TCCACGAGCGCCTTCTCCCTGAAGAACCTGGACAACTCACTGACCTGTGCACAACTCC 2046
QY 242 ValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMet 261
Db 2047 ATCCAGAACCTCTGGAACTCATGCCATCGGCATCCACTGCTTCGGCCAGACACATG 2106
QY 262 TrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTrpGly 281
Db 2107 TGGTCTAGCAGAGGTTCCAGGCCACCTGCAGAGATGGTGGTCCCAATGCTTGGTCC 2166
QY 282 SerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAsp 301
Db 2167 ACCATCATCTGCTGCGCATGAAGGATGCTGTGTCACACGACCTTCAGACCTCCCAAGGAC 2226
QY 302 HisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArg 321
Db 2227 ACGTGCAGTGTGCGAAGGCCAGCTTTGAGCTCTATGGCCGCTGACTTCGTGTCGGGGAG 2286
QY 322 AspPheArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThrPro 341
Db 2287 GACTTCAGCCCTGGCTGATTGAGATCAACGCCAGGCCCAACGATGGCACCTCCACAGCA 2346
QY 342 ValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrLysValAlaVal----- 359
Db 2347 GTCACTCCCGGCTCTGTGCTGGCGTCAAGCTGACACCCCTGCGCGTGGTCATGACCGG 2406
QY 360 -----AspArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValVal 377
Db 2407 AGGCTGACCGCAACTGTGACACGAGGACCTTTGAGCTCATCTATAGCAGCCTGTGTG 2466
QY 378 GluProProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgArgAla 397
Db 2467 GAGGTGCTCAATATATGTGGGCATCCGCTCTCTGTAGAGGCTTCACCATCAAGAACGCC 2526
QY 398 Arg-----ArgGlnValLeuProValCysAsnLeuLys 408
Db 2527 ATGCGCATGTGTCATCGCGGATGGGGTTCGCCGACGAGTCCCTCTGCTGACCCACCGA 2586
QY 409 AlaSerAlaSerLeu-LeuAspAlaGlnProLeuLysAla-ArgGlyProSerAlaMet- 427
Db 2587 GGCTCTGGGAGGCCGTGACCACTTCCCGAGCTCCACACGAGGAGGAGGAGGAGGAGGAGG 2646
QY 428 -----ProAspProAlaGlnGlyProProSerProAlaLeuGln----- 440
Db 2647 CCCCATGTACTCCGACACCGAGGCGAGTCTCTCAGACAGCAGACAGCAGCAGCAGCAG 2706
QY 441 -----ArgAspLeuGlyLeuLysGluGluLys-----GlyLeuP 452
Db 2707 ACTAAGGCCCTGTGACACAGGCAAGGCTTTGAGGACTCTACCCAGGCTTAAGGCTTTC 2766
QY 452 roLeuAlaLeuLeuAlaPro--LeuArgGlyAla-----AlaGluSerG 466
Db 2767 ATTTCCCTCCACCGAACCTTGATTTCAGGTGGCCACCGACATCTCTGAAGCCAAAGAG 2826
```

```
QY 466 lYgLYAlaAlaGlnProThrArgThr-----L 475
Db 2827 GTGGGCTCGACTGTGACTCACCCAGTGGACAGTGTGACACGGGTGAGGCTGG 2886
QY 475 ySAlaAlaGlyLysValGluLeuProAlaCysProCysArgHisValAspSerGlnAlaP 495
Db 2887 AGGGACAGCAGAGGAGGAGTCCACAGGCTGG----- 2918
QY 495 roAsnThrGlyValProValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsn- 514
Db 2919 -----CTGGCACCCCAAGGAGAGCTGGTCTCCCTCAGAAGCCCTT 2961
QY 515 -----AlaHisProLeuGluProValLeuArgGlyLeuLysThrAlaG 529
Db 2962 CTTCCACAGACTTCTGATCATCTCCCTCTTCTCCCTCTCT-----TCACCC 3009
QY 529 luGlyAlaLeuArgPro-----ProProGlyGly 538
Db 3010 GAGGCTCTGCTCTCTGTCCTCCGAGGCCCCAGCTGGA 3050

RESULT 4
AX301197 2380 bp DNA linear PAT 30-NOV-2001
LOCUS Sequence 39 from Patent W00185942.
DEFINITION AX301197
ACCESSION AX301197
VERSION AX301197.1 GI:17382228
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
1 Yue,H., Tang,Y.T., Au-Young,J., Lu,D.A., Baughn,M.R., Hillman,J.L.,
Azimzai,Y., Lal,P., Yao,M.G., Bandman,O., Burford,N., Batra,S.,
Kearney,L. and Pollicky,J.L.
Cytoskeleton-associated proteins
Patent: WO 0185942-A 39 15-NOV-2001;
Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
1..2380
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/notes="Incyte ID No: 2156553CB1"
ORIGIN

Alignment Scores: 4.05e-47 Length: 2380
Pred. No.: 1070.50 Matches: 250
Score: 54.8% Conservative: 81
Percent Similarity: 41.4% Mismatches: 198
Best Local Similarity: 37.5% Indels: 76
Query Match: 6 Gaps: 12
DB:

US-10-635-977-2 (1-541) x AX301197 (1-2380)
QY 2 AlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerArgSer 21
Db 285 GCCCGCAACGTTCTCAAGCTGGTGGTGAAGTCTGAG-----TGAAGTCA 329
QY 22 LysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerSerArgGlnAspAla 41
Db 330 TACCCTATTTCAGGCGAGTAGAGGAGGCGCTTCAGAGACAGCAGCAGCCAAAGAACAGGAG 389
QY 42 GluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLys 61
Db 390 AAAAACCCAGTGTG-----GTGTCCTCCAGAGTTTGTGGATGAGCTCTGTGT 437
QY 62 ValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAspAla 81
Db 438 GCGTGCAGGAGGTACCTTAGCAACTTGGCCCACTATGACATCGACAAGGACCTGGAGGCC 497
```

```

QY      82 ValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrTrpSerLeuVal 101
Db      498 CCGCTGTACTACCCCGAGGCGCTGGTCCCTCTTCTCCAGCGCTACTACCAAGTGGTC 557

QY      102 HisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeuLeu 121
Db      558 CACGAGGGGCGAAGTCTAGGACCTCGACACTCAGGTCCAGCGCTGTGAGGACATCTCG 617

QY      122 AsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIle 141
Db      618 CAGCAGCTGCAGCGCGTGTACCCAGATAGACATAGGAAGGGGATCGCAACATCTGGATC 677

QY      142 IleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGlu 161
Db      678 GTGAAGCCAGGAGCGCAAGTCCGTGGACGAGGACATCATGTGCATGACCACTCGAGGAG 737

QY      162 IleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnIleTrpValValGln 181
Db      738 ATGCTGAAGCTGGTGAACGGCAACCCCGTGGTGAAGGACGGCAAGTGGTGGTGCAG 797

QY      182 LysTrpIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPhe 201
Db      798 AAGTATATTGAGCGGCCCTCTCTCATCTTTGGCACCAGTTTGACCTCAGACAGTGGTTC 857

QY      202 LeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPhe 221
Db      858 CTGGTAACACTGACTGGAACCCCACTTACCGTGTGGTCTACCGGACAGCTATATCCGCTTT 917

QY      222 SerThrGlnArgPheSerLeuAspLysIleValAspSerAlaIleHisLeuCysAsnAla 241
Db      918 TCCAGCGACCCCTCTCCCTGAAGAACCTGGCAACTCAGTCACCTGTGCAACAACTCC 977

QY      242 ValGlnLysTrpLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMet 261
Db      978 ATCCAGAGCAGCTGGAGACTCATGCCATCGGCATCCACTGCTTCGCCACAGCAACATG 1037

QY      262 TrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTrpGly 281
Db      1038 TGGTCTAGCCAGAGTTCCAGGCCACCTGCAGGAGATGGGTGCCCAAACTCTTGGTCC 1097

QY      282 SerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAsp 301
Db      1098 ACCATCATCTGGCTGGCATGAAGGATGCTGTGATCCACGCACTTCAGACCTCCAGGAC 1157

QY      302 HisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArg 321
Db      1158 ACCGTGCAGTTCGGAAGCCAGCTTGGAGCTCTATGGCGTGAATCTTGTGTGGGAG 1217

QY      322 AspPheArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThrPro 341
Db      1218 GACTTCCAGCCCTGGCTGATTGAGATCAAGGCCAGCCACGATGGCACCTCCACAGCA 1277

QY      342 ValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaVal----- 359
Db      1278 GTCACTGCCCGCTCTGTCTGGCTGCAAGCTGACACCCCTCGCGGTGTCTATGACCGG 1337

QY      360 -----AspArgSerCysAspIleGlyAsnPheGluLeuLeuTyrArgGlnProValVal 377
Db      1338 ATGCTGGACCGCACTGTGACAGAGCCCTTTGAGCTCACTATAAGCAGCGCTGCTGTG 1397

QY      378 GluProProProSerGlySerAspLysCysValAlaGlyValSerValArgArg--- 396
Db      1398 GAGGTGCCTCAATATGTGGCATCCGGCTCTCTGGTAGAGGCTTCACCATCAAGAGGCC 1457

QY      397 -----AlaArgArgGlnValLeuProValCysAsnLeu----- 407
Db      1458 ATGGCGATGTGTCACTCGCGGATGGGGTCCGCCAGCAGTCCCTCTGCTGACCCAGCGA 1517

QY      408 -----LysAlaSer-AlaSerLeuLeuAspAlaGlnProLeu---LysAl 421
Db      1518 GGCTCTGGGGAAGCAAGGACTCGGGGATCCCTACCCACAGGTCACTGCTTAGGAAGGC 1577

```

```

QY      421 aArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln-- 440
Db      1578 ACTGGGGCCAGGAGCTGGGGGCACAGTGAGAGCCAGTCTCCACTGCCACCACTTCAGCC 1637

QY      441 -ArgAspLeuGlyLeuGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuLeu 460
Db      1638 CCCGAAAGGGGAAAGAAAGGCAAGGCGCAAGGCGCACAGCCCTGGTCTGCCCAATCTC 1697

QY      460 gGlyAlaAla-----GluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAl 477
Db      1698 TGGGAGTGGGATGCCCCCAGCAGCAGGATGGGCTGCATTTTC-ACCATGACCTTTCTAG 1756

QY      477 aGlyLysValGlu-----LeuProAlaCysProCysArgHisValAspSerGlnAl 494
Db      1757 TGGGACAGCGCAACCCCACTTGAACAGATTGCCACTGAGTCCGGAAGAACCCCAAGGC 1816

QY      494 aProAsnThrGlyValProValAlaGln-----ProAlaLys 506
Db      1817 CTGGGTAGACCAATTCCTCCCAAAACACCCGAGTGTTCAGGCGGATTTATTCCTGCTCT 1876

QY      506 sSerTrpAspProAsnGlnLeuAsnAlaHisPro----- 517
Db      1877 CCAGGCCCTCCCAACCACTGGATGAGCCACCCCAAGAGCCAGCAGTAGCAAGTA 1936

QY      517 ----- 517
Db      1937 AAAGCCACTACTCAAAAGTATTGTTTAAANAATACACAGCCAAATAGCTGGCGACGGTG 1996

QY      518 -LeuGluProValLeuArgGlyLeuLysThrAlaGlyAlaLeuArgProProGln 537
Db      1997 GTGTGAGCTGTGGTCCAGGCTACTCGCGAGCTTAATGAGGAT--CGCTTGAGCCCGGG 2053

QY      537 yGlyLysGly 540
Db      2054 AGGTCAAGGC 2063

RESULT 5
LOCUS      BC098298
DEFINITION Homo sapiens tubulin tyrosine ligase-like family, member 3, transcript variant 1, mRNA (cDNA clone MGC:120531 IMAGE:40025661), complete cds.
ACCESSION BC098298
VERSION      BC098298.1 GI:68532483
KEYWORDS      MGC.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 2097)
AUTHORS      Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullen, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalusz, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Mammalian Gene Collection Program Team
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

```


Db 997 ATGCTGAAGCTGGTGAACGCAACCCGCTGGTGTGATGAAGCAGCGCAAGTGGGTGGTGACG 1056
 Qy 182 LysTyrlleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPhe 201
 Db 1057 AAGTATATATGAGCGGCCCTCTCTCATCTTTGGCACCAAGTTTGACCTCAGACAGTGGTTTC 1116
 Qy 202 LeuValThrAspTrpAsnProLeuThrIleTTPheTyrlsGluSerTyrlsLeuArgPhe 221
 Db 1117 CTGGTAACTGACTGGAAACCACTTACCGTGTGTGTTCTACCGGACAGCTATATCCGCTTT 1176
 Qy 222 SerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAla 241
 Db 1177 TCCAGCAGCCCTCTCTCCCTGAAGAACCTGGACAACTCAGTCAGCTGTGCAACACTCC 1236
 Qy 242 ValGlnLysTyrlsLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMet 261
 Db 1237 ATCCAGAAGCACTGGGAACCTCATGCCATCGGCATCTCCACTGCTTCGCGCAGACAATG 1296
 Qy 262 TrpThrSerThrArgPheGlnGluTyrlsLeuGlnArgGlnGlyArgGlyAlaValTrpGly 281
 Db 1297 TGGTCTAGCCAGAGGTTCAGGCCCACTGCGAGGAGATGGGTGCCCAATGCTTGGTCC 1356
 Qy 282 SerValIleTyrlsProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAsp 301
 Db 1357 ACCATCATCTGCTGGCATGAGGATGCTGTGATCCAGCACTTCAGACCTCCAGGAC 1416
 Qy 302 HisValGluProArgLysAsnSerPheGluLeuTyrlsGlyAlaAspPheValLeuGlyArg 321
 Db 1417 ACCGTGCACTGCGAAGGCCAGCTTGTAGCTCTATGGCGCTGACTGCTGTGTTCCGGGAG 1476
 Qy 322 AspPheArgProTrpIleGluLeuAsnSerSerProThrMetHisProSerThrPro 341
 Db 1477 GACTTCCAGCCCTGCTGATGAGATCAACGCCAGCCCAACATGGCACTCCACAGCA 1536
 Qy 342 ValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaVal 359
 Db 1537 GTCACTGCGCGCTCTGTGCTGGCTGACGCTGACCCCTGGCTGATTCACCGG 1596
 Qy 360 -----AspArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValVal 377
 Db 1597 ATGCTGGACCGCACTGTGACAGGAGCTTTGAGCTCATATGAAGCAGCTCTGCTG 1656
 Qy 378 GluProProPheSerGlySerAspLysCysValAlaGlyValSerValArgArgAla 397
 Db 1657 GAGTGGCTCAATATGTGGGCACTCGCGCTCTGCTGGTAGAGGCTTCACCATCAAGAAG 1713
 Qy 398 ArgArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGln 417
 Db 1713 ----- 1713
 Qy 418 ProLeuLysAlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerPro 437
 Db 1714 -----CCCATGGCGATG----- 1725
 Qy 438 AlaLeuGlnArgAspLeuGlyLeuLysGluLysGlyLeuProLeuAlaLeuLeuAla 457
 Db 1726 ---TGTCTATCGCGATGGGTGCGC-----CCAGCAGTCCCTCTGCTG 1767
 Qy 458 ProLeuArgGlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgTrpLysAlaAla 477
 Db 1768 ACCCAGCGAGGCTCTGGGGAAGCCGCTGCCACTTCCCGAGCTCCACCAAGGCC--- 1824
 Qy 478 GlyLysValGluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThr 497
 Db 1825 -----CAGTGGCTTCTCC---CCATGTACTCCGACACAGGCGGCTCTCAGACG 1874
 Qy 498 GlyValProValAlaGlnProAlaLysSerTrp-----AspProAsnGln--- 512
 Db 1875 -----ACAGCACACAGAGCTGGTGGGCACTAAGGCCCTGTGACACACAGGCAAGGC 1925
 Qy 513 -----LeuAsnAlaHisProLeuGluProValLeuArg 523
 Db 1926 CTTGAGGACTACCCAGCGGTAAAGTCTTCATTTTCCCTCCCAACACCTTGATTTCAA 1985

Qy 524 Gly-----LeuLysThrAlaGluGlyAlaLeuArgPro-----Pro 535
 Db 1986 GGTGGCACCAGCATCTCTGAAGCAAGAGGCTCTGCTCTCTGCTCGAGGCC 2045

Qy 536 ProGlyGly 538
 Db 2046 CCAGCTGGA 2054

RESULT 8

AK023960

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

construction, 5'- & 3'-end one pass sequencing and clone selection:
 Helix Research Institute (supported by Japan Key Technology Center
 etc.) and Department of Virology, Institute of Medical Science,
 University of Tokyo.

FEATURES

source

Location/Qualifiers

1..2326

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="THYRO1001738"

/tissue_type="thyroid gland"

/clone_lib="THYRO1"

/note="cloning vector: pME18SFL3"

94..>2326

/notes="unnamed protein product"

/codon_start=1

/protein_id="BAB14741.1"

/db_xref="GI:10436084"

/translation="MNRLNRAKIYVERAVKQKIFTIOGYPVIRCLRRRGWVEKKM
 VHSGLTLPQKDLSSAMGSDTTEDEDEDEFEQPSQSFDFDLDLKFDDLDGTH
 ALMGVGLNLRLNLPWFEDVDANSFFPRCYCLGAEDDKAFIEDFWLTAARNVLKLVVK
 SEWKSYPFOIAVEEASGDKPKQKRNPLVSPFDEALCAACEEYLSLNAHMDIDKD
 LEAPLYTPEGSLFLQRYQVQVHEGAELRLDTPQRCEDILQQLQAVVPQIDMEGD
 RAIWVKGAKSRGRCIMCMHLEMLKLAVNGNPVVMKQKVVOKYIERPLLIFGTK
 FDLRQWFLVTDNPLTVWYPRDSYIRFSTQPSLKNLDSVHLKNNISIOKHLNENCHR
 HPLLPDNNWSSQRFQAHQEMGAPNAWSTIIIVPGMKDAVIHALQTSQTVQCRKASF
 ELYGADFVEGEQFQWLIIEINAPSTWAPSTAVTARLCAGVQADTLRVLDLRNCD
 TGAFLIYKQPAVEYVQVIGIRLLVEGFTIKKPMANCHRMGMVRPAVPLITQMGSEA
 RHPLPSLHTKAQPSLPHVLRHOGVLRROHSLKLVGTAKLSTTGKFLAPITAKVFSL
 PNULDPKVAPSILKPKAPALLCLRGPLEVPCCLPLKSEQFLAPVGRSRPKANSRP
 DCDKPRAEACPMKRLSLPLPLPLVGTGFRRRGLGDMKLGKPLLRFTALVLDPTNKK
 KOVK"

ORIGIN

Alignment Scores:

Pred. No.: 2,88e-46 Length: 2326
 Score: 1054.00 Matches: 233
 Percent Similarity: 55.4% Conservative: 79
 Best Local Similarity: 41.4% Mismatches: 166
 Query Match: 36.9% Indels: 86
 DB: 8 Gaps: 12

US-10-635-977-2 (1-541) x AK023960 (1-2326)

Qy 2 AlasSerSerLeuLysTrpValSerHisGlnSerCysSerArgSerSerArgSer 21
 Db 544 GCCCGCAACGTTCTCAAGCTGGTGGTGAAGTCTGAG-----TGAAGATCA 588
 Qy 22 LysProArgaspGlnArgGluGluAlaGlySerSerAspLeuSerSerArgGlnAspAla 41
 Db 589 TACCCTATTACGGCAGTAGAGGAAGAGCGCTCAGGAGCAAGCAGCAGCAAGAACAGGAG 648
 Qy 42 GluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLys 61
 Db 649 AAAAACCCAGTGTG-----GTGTCCCGCAGAGTTGTGGATGAAGCTCTGTGT 696
 Qy 62 ValCysGlnAlaTrpLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAspAla 81
 Db 697 GCGTGCAGGAGTACCTTAGCAACTTGGCCACATGACATCGACAAAGGACCTGGAGGCC 756
 Qy 82 ValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTrpTyrSerLeuVal 101
 Db 757 CCGCTGTACCTCACCCCGAGGGCTGTGCTCTTCTCCAGCGCTACTACCAAGTGGTC 816
 Qy 102 HisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeuLeu 121
 Db 817 CACGAAGGGCAGAACTCAGCACCTCGACACTCAGGTCCAGCGCTGTGAGGACATCTGT 876
 Qy 122 AsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIle 141
 Db 877 CAGCAGCTGCAGCGCGTGGTACCCACAGATAGACATGGAAGGGGATCGCAACATCTGGATC 936
 Qy 142 IleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGlu 161

Db 937 GTGAAGCCAGGAGCCAGCAAGTCCCGCGGAGCGAGGCATCATGTGATGCACCACTGGAGGAG 996
 Qy 162 IleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValGln 181
 Db 997 ATGCTGAAGCTGGTGAACCGCAACCCCGTGGTGTATGAAGGACGGCAAGTGGGTGGTGCAG 1056
 Qy 182 LysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPhe 201
 Db 1057 AAGTATATTAGCGGCCCTCTCTCATCTTTGGCACCAAGTTTGACTCAGACAGTGGTTC 1116
 Qy 202 LeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPhe 221
 Db 1117 CTGGTAACCTGACTGGAACCCACTACCGTGTGGTCTACCGCGACAGCTATATCCGCTTT 1176
 Qy 222 SerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAla 241
 Db 1177 TCCAGCGACCCCTTCTCCCTGAAGAACCTGGCAACTAGTGCACCTGTGCAACAACCTCC 1236
 Qy 242 ValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMet 261
 Db 1237 ATCCAGAACCACTGGAGAACTCATGCCATCGGCATCCACTGCTTCGGCCAGACACAATG 1296
 Qy 262 TrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTrpGly 281
 Db 1297 TGGTCTAGCCAGAGGTTCCAGGCCACCTGCAGGAGATGGGTGCCCAATGCTGTGCTCC 1356
 Qy 282 SerValIleTyrProSerMetLysAlaIleAlaHisAlaMetLysValAlaGlnAsp 301
 Db 1357 ACCATCATCGTCTGGCATGAAGATGCTGTATCCACGCCACTTCAGACCTCCCAAGGAC 1416
 Qy 302 HisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArg 321
 Db 1417 ACCGTGACGTGTCGAAGGCCAGCTTTGAGCTCTATGGCGCTGACTTCGTGTGCGGGAG 1476
 Qy 322 AspPheArgProTrpLeuIleGluLeuAsnSerSerProThrMetHisProSerThrPro 341
 Db 1477 GACTTCCAGCCCTGGCTGATTGAGATCAACGCCAGCCCCACGATGGCACCTCCACAGCA 1536
 Qy 342 ValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaVal----- 359
 Db 1537 GTCACCTGCCCGCTCTGTGCTGGCGTGAAGCTGACACCTGGCGGTGTCATGACCCGG 1596
 Qy 360 -----AspArgSerCysAspIleGlyAsnPheGluLeuTrpArgGlnProValVal 377
 Db 1597 ATGCTGACCGCAACTGTGACAGAGGCTTTGAGCTTATCTATAGCAGCCTGCTGTG 1656
 Qy 378 GluProProProSerGlySerAspLeuCysValAlaGlyValSerValArgAla 397
 Db 1657 GAGTGCTCAATATGTGGGCATCCGCTCTCTGTAGAGGGCTTCACCATCAAGAAG--- 1713
 Qy 398 ArgArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGln 417
 Db 1713 ----- 1713
 Qy 418 ProLeuLysAlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerPro 437
 Db 1714 -----CCATGGCGATG----- 1725
 Qy 438 AlaLeuGlnArgAspLeuGlyLeuLysGluLysGlyLeuProLeuAlaLeuLeuAla 457
 Db 1726 ---TGTCATCGCGGATGGGGTCCGC-----CCAGCAGCTCCTCTGCTG 1767
 Qy 458 ProLeuArgGlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAla 477
 Db 1768 ACCCAGCGAGGCTCTGGGGAAGCCCGTCAACCTCCCGACCTCCACCAAGGCC--- 1824
 Qy 478 GlyLysValGluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThr 497
 Db 1825 -----CAGCTGCTCTTCTCC--CCATGTACTCCGACACCGAGGCCGCTCTCAGACG 1874
 Qy 498 GlyValProValAlaGlnProAlaLysSerTrp-----AspProAsnGln--- 512

```

Db 1875 -----ACAGCAGCAGCAAGCTGTGGGCACTAAGCCCTGTGTCACACAGGCAAGGC 1925
QY 513 -----LeuAsnAlaHisProLeuGluProValLeuArg 523
Db 1926 CTTGAGGACTTACCCACGCTCAAGGCTCTTCAATTCCTCCACCAACCTTGATTCAA 1985
QY 524 GY-----LeuysThrAlaGluGlyAlaLeuArgPro-----Pro 535
Db 1986 GGTGGCACCAGCATCTGAAGCAAGAAAGGCTCTGCTCTCTGTGCTCCGAGGCC 2045
QY 536 ProGlyGly 538
Db 2046 CCAGCTGGA 2054

RESULT 9
HSM800637
LOCUS
DEFINITION Homo sapiens mRNA; cDNA DKFp434B103 (from clone DKFp434B103);
partial cds.
ACCESSION AL096725
VERSION AL096725.1 GI:5419858
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 2553)
Wambutt, R., Heubner, D., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
Direct Submission
Submitted (16-JUN-1999) MIPS, Am Klopferspitze 18a, D-82152
Martinsried, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFp434B103) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.

FEATURES
source
1..2553
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFp434B103"
/db_xref="taxon:9606"
/clone="DKFp434B103"
/tissue_type="testis"
/clone_lib="434 (synonym: htes3). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
1..2553
/gene="DKFp434B103"
921..1950
/gene="DKFp434B103"
/feature="similarity to tubulin-tyrosine ligase"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB46375.1"
/db_xref="GI:5419859"
/db_xref="GOA:Q9Y4R7"
/db_xref="UniProt/Swiss-Prot:Q9Y4R7"
/translation="MDIKLEAPLTPGMSLFLQRYQVHVHEGARLHLDQVOR
CEDILQOLQAVVPOIDMEGDRNIWIKPKAKSRGGINCMDBLEMLKLVNPNVVK
DKWVVKXIERPLIFGTFKFDLQWFLVDWNLPTVWFYRDSYRSTQFNPKLND
NSVHLCKNSIKHLENSCHRPFLPLPPDNWSSQFQALHQMGNAPNASTIIVFGMKD
AVIALQTSQDQVCRKASFLYGADVFVGFQFLEINASPTMAPSTAVTARLCA
GVQADTLRWIDRLMDRNCDTGAFELIYKQVPTTSPASTPRPSPCLLPMYSOTRARSDD
DSTASWALRPCRQAPR"
2511..2516
polyA_signal
/gene="DKFp434B103"
2528
polyA_site

```

/gene="DKFp434B103"

ORIGIN

```

Alignment Scores:
Pred. No.: 1,91e-44 Length: 2553
Score: 1020.00 Matches: 220
Percent Similarity: 55.5% Conservatives: 69
Best Local Similarity: 42.2% Mismatches: 151
Query Match: 35.7% Indels: 81
DB: 8 Gaps: 12

US-10-635-977-2 (1-541) x HSM800637 (1-2553)
QY 54 GlnLeuValAspPileAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGlu 73
Db 835 GAGTTTGTGGATGAAGCTCTGTGTGTCGCGAGGAGTACTTTAGCAACTTGGCCACATG 894
QY 74 AspPileAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu 93
Db 895 GACATCGACAAGGACCTCGAGGCGCCCTGTGTACCTCACCCCGAGGGCTGGTCTCTTC 954
QY 94 ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
Db 955 CTCGAGCGCTACTACAAAGTGTCCACGAAGGGGAGAACTCAGGCACCTCGACACTCAG 1014
QY 114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspPile 133
Db 1015 GTCCAGCGCTGTGAGGACATCTCTGCAGGAGCTGCAGGCGCGTGTATCCCGAGACATG 1074
QY 134 AspGlyLeuArgAsnIleThrPileIleLysProAlaAlaLysSerArgGlyArgAspPile 153
Db 1075 GAAAGGGATCGCAACATCTGGATCTGTAAGCAGGAGGCAAGTCCCGCGAGCGAGCATC 1134
QY 154 ValCysMetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLeuSer 173
Db 1135 ATGTGCATGGACCACTCGAGGAGATGCTGAAGCTGGTGAAGCGCAACCCCGTGGTGATG 1194
QY 174 ArgAspAsnLysTrpValGlnLysTyrIleGluThrProLeuLeuIleCysAspThr 193
Db 1195 AAGGACGCAAGTGGTGTGTCAGAGAATATATTGAGCGGCCCTCTCTCATCTTTGGCACC 1254
QY 194 LysPheAspPileArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
Db 1255 AAGTTTGTACCTCAGACAGTGTTCCTGTGTAACCTGGAACCCACTTACCGCTGGTTC 1314
QY 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
Db 1315 TACCGGACAGTATATCCGCTTTTCCAGCGAGCCCTTCTCCCTGAAGAACCTGGACAAAC 1374
QY 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
Db 1375 TCAGTGACCTGTGCAACAACCTCCATCCAGAGCACCTGGAGAACTCATGCGCATCGGCAT 1434
QY 254 ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArg 273
Db 1435 CCAGTGTCTCCGCCAGACACATGTGTGTCTAGCCAGAGAGTTCACAGCGCCACCTGCAGGAG 1494
QY 274 GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla 293
Db 1495 ATGGTGTGCCCCAATGTGTGTTCACCATCATCTGTGCTGGCATGAAGATGCTGTGATC 1554
QY 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
Db 1555 CACGCACTTCAGACCTCCAGGACACCGTGCAGTGTCCGAGGCCAGCTTTGAGCTCTAT 1614
QY 314 GlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluIleAsnSerSer 333
Db 1615 GCGCTGACTTCGTGTTCGGGGAGGACTTCCAGCGCTGGCTGATTTAGATCAACGCCAGC 1674
QY 334 ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353
Db 1675 CCCAGGATGGCACCCCTCCACAGCAGTCACTCCCGGCTCTGTGCTGGCGTGCAGAGTGCAT 1734

```


QY 388 sValAlaGlyValSerValArgAlaArgGlnValLeuProValCysAsnLeuIy 408
 Db 2312 ----- 2312
 QY 408 sAlaSerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArgGlyProSerAlaMetPr 428
 Db 2313 ----CCAGTGCCTTCTCCCTGACTCGACACAGGCGCAGGCTCTCAGACGACAC 2368
 QY 428 oAspProAlaGlnGlyProProSerProAlaLeuGlnArgAspLeuGlyLeuLysGluGl 448
 Db 2369 ACAGCAAGCTGGTGGCACTAAGGCCCTGTGACACACAGCAAA---GGCTTGAAGACTC 2425
 QY 448 uIys-----GlyLeuProLeuAlaLeuAlaPro---LeuArgGlyAla 462
 Db 2426 TACCCACGCTAAGGCTTCTTCATTTCCCTCCACCGAACCTGATTTCAAGGTGGCACCA 2485
 QY 463 -----AlaGluSerGlyGlyAlaAlaGlnProThrArgThr----- 474
 Db 2486 GCATCTGAAGCCAGAAAGGTGGCTCGACTGTGACTCACACCCAGTGCACAGTGTCT 2545
 QY 475 -----LysAlaAlaGlyLysValGluLeuProAlaCysProCysAr 488
 Db 2546 GAGCAGCGGCTCAGGCGTGGAGGCACAGGAGGCGAGCTCCAGGCTGG----- 2597
 QY 488 gHisValAepSerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTr 508
 Db 2598 -----CTGGCACCCCAAGGAGAGAGCTG 2620
 QY 508 pAspProAsnGlnLeuAsn-----AlaHisProLeuGluProValLe 522
 Db 2621 GTCTCCCTCAGACGCGCTCTCTCCACAGACTTCTGATCATCTCTCTCTCTCTCTCT 2680
 QY 522 uArgGlyLeuLysThrAlaGluGlyAlaLeuArgPro-----ProProGlyGl 538
 Db 2681 T-----TCACACGAGGCTCTGCTCTCTGCTCGAGGCCCGAGCTGG 2728
 QY 538 Y 538
 Db 2729 A 2729
 RESULT 11
 LOCUS BC099735
 DEFINITION Homo sapiens tubulin tyrosine ligase-like family, member 3, mRNA
 ACCESSION BC099735
 VERSION BC099735.1 GI:71043427
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1769)
 AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
 Klausner,R.D., Collins,F.S., Wagner,L.H., Shenmen,C.M., Schuler,G.D.,
 Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
 Hopkins,R.P., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
 Datchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
 Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
 Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
 Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
 Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
 Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
 Fahey,J., Heiton,E., Kettelman,M., Madan,A., Rodrigues,S.,
 Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
 Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
 Schnurch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

CONSRMT
TITLE

Mammalian Gene Collection Program Team
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

REFERENCE
2 (bases 1 to 1769)AUTHORS
CONSRMT
TITLE

NIH MGC Project
 Direct Submission
 Submitted (15-JUL-2005) National Institutes of Health, Mammalian
 Gene Collection (MGC), Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 CONTACT: MGC help desk
 Email: cgapbs@mail.nih.gov

REMARK
COMMENT

Tissue Procurement: Baylor Human Genome Sequencing Center
 cDNA Library Preparation: Baylor Human Genome Sequencing Center
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@hgrl.nih.gov
 Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
 Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
 Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
 Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Lalic,P., Legaspi,R.,
 Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
 McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
 Tsurgueon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
 Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAM Plate: 4 Row: p Column: 10.

FEATURES
source

1..1769
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:120532 IMAGE:40025662"
 /tissue_type="PCR rescued clones"
 /clone_lib="NIH MGC_283"
 /note="vector: PCR-Blunt II-TOPO with reversed insert;
 clone identification sequence tag: AGTCAG"
 1..1769
 /gene="TTLL3"
 /note="synonyms: DKFP434B103, HOTTL"
 /db_xref="GeneID:26140"
 360..1418
 /gene="TTLL3"
 /codon_start=1
 /product="TTLL3 protein"
 /protein_id="AAH99735.1"
 /db_xref="GI:71043428"
 /db_xref="GeneID:26140"
 /translation="MDIDKDLAEPLYPTEGMSLFQRYQVWHEGAEHLDTQVOR
 CEDTLQLOAVVPOIDMEGDRNIWVKGAKSRGIMCMOHLKMLVNGNPVVMK
 DKGWVQYKIERPLLIIGTRFDLQWFLYTDWNPULTWYFDRDSYRSTQFSLNKLD
 NSVHLKNKSIQKLENSCHRHPLPPDNWSSQRFQALHQEMGAPNAWSTIIVPGMKD
 AVTHALQSDTQVCRKASFLYGDVFGEDFQFLIEINASPTMAEPSTAVTARLCA
 GVQADTLRVTDVDRNCDTGAFELIYKQPVTTSPASTPRPSCILLPMYSIDTRRSSD
 DSTASWALRPCRQARP"

gene

CDS

ORIGIN

Alignment Scores:

Pred. NO.: 1.19e-43 Length: 1769
 Score: 1001.50 Matches: 218
 Percent Similarity: 56.9% Conservative: 65
 Best Local Similarity: 43.9% Mismatches: 142
 Query Match: 35.1% Indels: 74
 DB: 8 Gaps: 10

US-10-635-977-2 (1-541) x BC099735 (1-1769)

QY	54	GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrIleuGlyGlnLeuGluHisGlu	73	QY	393	SerValArgAlaArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSer	412
DB	303	GAGTTTGTGATGAAGCTCTGTGTGGTGGAGAGTACCTTAGCAACTTTGGCCACATG	362	DB	1383	GCACTAAGCCCTGTGCACACGACGAGCCTTG-AGGACTCTACCCAGCGTAAAGTTC	1441
QY	74	AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu	93	QY	413	LeuLeuAspAlaGlnPro-----LeuLysAlaArg	422
DB	363	GACATCGACAGACCTGGAGGCGCCGCTGTACTCACCCTCCGAGGCTGTCCCTCTTC	422	DB	1442	TTTCATTTCCCTCCCAACCACTTGATTTCAAGGTGGCACCAGCATCTCTGAAGCCAAGA	1501
QY	94	ThrGlnGlnTyrTrpSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr	113	QY	423	GlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGlnArgAsp	442
DB	423	CTCCAGCGTACTACCAAGTGGTCCAGGAGGGCAGAACTCAGGCACCTCGACACTCAG	482	DB	1502	AA-GGCTCTGTCTCTCTCGAGGCCCCAGCTGGAGTGCCTTG-----	1551
QY	114	PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle	133	QY	443	LeuGlyLeuLysGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuArgGlyAla	462
DB	483	GTCAGCGCTGTGAGGACATCTGCAGACAGCTCGAGCGCGTGTACCCAGATAGACATG	542	DB	1552	-----TTGCCTCTG-----CCC	1563
QY	134	AspGlyLeuArgAsnIleTrrIleIleLysProAlaAlaLysSerArgGlyArgAspIle	153	QY	463	AlaGluSerGlyGlyAlaAlaGlnProThrArg---ThrLysAlaAlaGlyLysValGlu	481
DB	543	GAAGGGGATCGCAACATCTGGATCGTGAAGCCAGGAGCCAAAGTCCCGTGGACAGGCATC	602	DB	1564	TCTGAAGTCGGAACAATTCTAGCACCTGTGCGAAGTCAAGGCCCAAGCAAAATTCAG	1623
QY	154	ValCysMetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLeuSer	173	QY	482	-----LeuProAlaCysProCysArgHisValAsp	491
DB	603	ATGTGCATGACACCTGGAGGAGATCTGAAGCTGTGTGAAGCGCAACCCGCTGGTGATG	662	DB	1624	GCCAGACTGTGACAAACCCAGGCGTGAAGCCTGCC-----CATGAAGAG	1668
QY	174	ArgAspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThr	193	QY	492	SerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTrp	508
DB	663	AAGGACGGCAAGTGGGTGGTGCAGAAATATATTGAGCGCGCCCTCTCATCTTTGGCACC	722	DB	1669	GCTGAGCCCCCTGAAACC-----CCTGCCCTGTGTGG	1701
QY	194	LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe	213	RESULT 12			
DB	723	AACTTTGACCTCAGACAGTGTCTCTGGTAACCTGAACTGGAACCCACTTACCGTGTGGTTC	782	BC098361	1743 bp	mRNA	linear PRI 03-AUG-2005
QY	214	TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer	233	LOCUS	BC098361		
DB	783	TACCGCAGACTATATCCGCTTTCCACGACGCCCTTCTCCCTGGAAGAACCTGGACAAC	842	DEFINITION	BC098361.1	GI:68532603	
QY	234	AlaIleHisLysCysAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer	253	ACCESSION			
DB	843	TCAGTCACCTGTGCACAACTCCATCCAGAACACCTGGAGAACTCATCCCATCGGCAT	902	VERSION			
QY	254	ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArg	273	KEYWORDS			
DB	903	CCACTGCTTCGCCAGACAACATGTGTCTAGCCAGAGGTTCAGGCCACCTTCAGGAG	962	SOURCE			
QY	274	GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla	293	ORGANISM			
DB	963	ATGGGTGCCCAATGCTTGTCCACCATCATGCTGCTGCATGAAGGATGCTGTGATC	1022	REFERENCE			
QY	294	HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr	313	AUTHORS			
DB	1023	CAGGCATCTCAGACCTCCAGGACACCGTGCATGTGCGAAGGCCAGCTTTGAGCTCAT	1082				
QY	314	GlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluIleAsnSerSer	333				
DB	1083	GGCGCTCACTTGGTTCGGGGAGACTTCAGCCCTGGCTGTGATGATGATCAAGCCGAGC	1142				
QY	334	ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp	353	CONSRSTM			
DB	1143	CCCACGATGGCACCTCCACAGCAGTCACTGCCCGGCTGTGTCGCGTGAAGCTGAC	1202	TITLE			
QY	354	ThrIleLysValAlaVal-----AspArgSerCysAspIleGlyAsnPheGlu	369	JOURNAL			
DB	1203	ACCTGCGCGTGGTCAATTGATCCGCGAGGCTGGACCGCAACTGTGACAGGAGCCTTTGAG	1262	PUBMED			
QY	370	LeuLeuTrpArgGlnProValValGluProPro-----	380	AUTHORS			
DB	1263	CTCATCTATAAGACGCCCGTCACCACTTCCCGCCAGCTCCACCAAGGCCCGAGCTGCCTT	1322	CONSRSTM			
QY	381	---ProPhe-----SerGlySerAspLeuCysValAlaGlyVal	392	JOURNAL			
DB	1323	CTCCCCATGTACTCCGACACCAAGGCCGAGGCTCTCTCAGACGACACGACAGCAAGCTGTGG	1382	REMARK			
				COMMENT			

BC098361 1743 bp mRNA linear PRI 03-AUG-2005
Homo sapiens tubulin tyrosine ligase-like family, member 3, mRNA
(CDNA clone MGC:120529 IMAGE:40025659), complete cds.

BC098361
MGC.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 1743)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Schetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1743)

NIH MGC Project
Direct Submission
Submitted (24-JUN-2005) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk

Email: cgabs-r@mail.nih.gov
 Tissue Procurement: Baylor Human Genome Sequencing Center
 cDNA Library Preparation: Baylor Human Genome Sequencing Center
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@hgrl.nih.gov
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
 Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
 Young, A., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAM Plate: 4 Row: 0 Column: 9.

FEATURES

source

1. 1743

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:120529 IMAGE:40025659"

/tissue_type="PCR rescued clones"

/clone_lib="NIH MGC 283"

/note="Vector: PCR-Blunt II-TOPO with reversed insert;

Clone identification sequence tag: GAAGCCCTGCTTCG"

1. 1743

/gene="TTLL3"

/note="synonyms: DKFZp434B103, HOTTL"

/db_xref="GeneID:26140"

360..1418

/gene="TTLL3"

/codon_start=1

/product="TTLL3 protein"

/protein_id="AAH98361.1"

/db_xref="GI:68532604"

/db_xref="GeneID:26140"

/translation="MDIDKLEAPLYLTPEGWSLFLQRYQVVEHGAELRLDTPQVR
 CEDLIQQAIVPQIDMGDRINWIKFGAKSRGRCINCMHLEMLKLVNGNPNVWK
 DGVVQYKIERPLIFGTQFLQWFLVTDWNPFLVYFDRSTQPSFLNKLD
 NSVHLNNSIOKHLISCHRPDPNDWSSORFOAHLQBMGAPNAWSTIIVPMKD
 AVIHALQTSQTVOCRKASFLYGLADFVEGDFQPMLEINASTPMAPSTAVTARLCA
 GYQADTLAVDIDRLDRNCDTGAPELYIKQPVITTSPTSPRPSCLLPYSIDTRSSD
 DSTASWALRPCRPOARD"

ORIGIN

Alignment Scores:

Pred. No.:	1,32e-43	Length:	1743
Score:	1000.50	Matches:	218
Percent Similarity:	56.7%	Conservative:	64
Best Local Similarity:	43.9%	Mismatches:	143
Query Match:	35.1%	Indels:	74
DB:	8	Gaps:	10

US-10-635-977-2 (1-541) x BC098361 (1-1743)

Qy	54	GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGlu	73
Db	303	GAGTTTGGATGAAGCTCTGTGCGTGCAGGAGTACTTAGCACTTGGCCCAATG	362
Qy	74	AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu	93
Db	363	GACATTCGACAGGACCTGGAGGCGCGCTGTACCTCACCCCGAGGGGTGGTCCCTTC	422
Qy	94	ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr	113
Db	423	CTCCAGCGCTACTACCAAGTGGTCCACGAGGGGCGAAGCTCAGGCACCTCGACACTCAG	482

Qy	114	PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle	133
Db	483	GTCCAGCGCTGTGAGGACATCTCTGCAGCAGCTGCAGGCGTGGTATAGACATG	542
Qy	134	AspGlyLeuArgAsnIleTrpIleLeuLysProAlaAlaLysSerArgGlyArgAspIle	153
Db	543	GAAGGGGATCGCAACATCTGGATCGTGAAGCCAGGAGCCAGTCCCGTGGACGAGCATC	602
Qy	154	ValCysMetAspArgValGluGluLeuGluLeuGluLeuAlaAlaAspHisProLeuSer	173
Db	603	ATGTGCATGGACCACTGGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGGATG	662
Qy	174	ArgAspAsnLysTrpValGlnLysTyrIleGluThrProLeuLeuLeuLysCysAspThr	193
Db	663	AAGGACCGCAAGTGGTGGTGCAGAAATATATTGAGCGGCCCTCTCTCATCTTTGSCACC	722
Qy	194	LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe	213
Db	723	AAGTTTGACCTCAGACAGTGGTTCCTGTTAACTGATCGAACCCATTAACCGTGGTTC	782
Qy	214	TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer	233
Db	783	TACCGCAGAGTATATCGCTTTTCCAGCAGGCCCTTCTCCCTGAAGAACCTGGACAAAC	842
Qy	234	AlaIleHisLysCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer	253
Db	843	TCAGTGCACTGTGCAACAACTCCATCCAGAGCACCTGGAGAACTCATGCGCATCGGCAT	902
Qy	254	ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArg	273
Db	903	CCACTGCTCCGCCAGACACATGTGTGTAGCCAGAGGTTCCAGGCCACCTGCAGGAG	962
Qy	274	GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla	293
Db	963	ATGGTGCCCAAAATGCTTGTGCCATCATCTGCTGCGCTGGCATGAAGATCTGTGATC	1022
Qy	294	HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr	313
Db	1023	CAGCACTTCAGACTCCCGCAGGACACCGTGCATATCTCGAAGGCCAGCTTTGAGCTCTAT	1082
Qy	314	GlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluIleAsnSerSer	333
Db	1083	GGCGCTGACTTCTGTTCTGGGGAGGACTTCCAGCCCTGGCTGATTTAGATCAACGCCAGC	1142
Qy	334	ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp	353
Db	1143	CCCAGATGGCACCTCCACAGAGTCACTCCCGGCTCTGTGCTGGCGGTCAAGCTGAC	1202
Qy	354	ThrIleLysValAlaVal-----AspArgSerCysAspIleGlyAsnPheGlu	369
Db	1203	ACCTGGCGGTGGTCAATTGACCGGAGGCTGGACCGCAACTGTGACACAGGAGCCTTGAG	1262
Qy	370	LeuLeuTrpArgGlnProValValGluProPro-----	380
Db	1263	CTCATCTATAAGCAGCGCGTCCACACTTCCCGCAGCCTCCACCAAGGCCAGCTGCGCTT	1322
Qy	381	---ProPhe-----SerGlySerAspLeuCysValAlaGlyVal	392
Db	1323	CTCCCATGTACTCCGACACCGAGGCGCGTCTTCAGACGACAGACAGCAAGCTGGTGG	1382
Qy	393	SerValArgArgAlaArgArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSer	412
Db	1383	GCATAAGGCCCTGTGCACACAGGCAAGGCTTG-AGGACTTCTACCCACGGCTAAGGTC	1441
Qy	413	LeuLeuAspAlaGlnPro-----LeuLysAlaArg	422
Db	1442	TTCAATTTCCCTCCACCGAACCTTGATTTCAAGGTGGCACCCAGCATCTCTGAAGCAAGA	1501
Qy	423	GlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGlnArgAsp	442
Db	1502	AA-GGCTCTCTGCTCTCTGCTCTCCAGGCGCCCGAGCTGGAAGTGGCTTG-----	1551
Qy	443	LeuGlyLeuLysGluGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuArgGlyAla	462

Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hirooka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yosida, M., Houta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Musahino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hiroo, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigetani, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togaehi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs

Nat. Genet. 36 (1), 40-45 (2004)
JOURNAL
PUBMED
REFERENCE
AUTHORS

14702039
2

Ninomiya, K., Wagatsuma, M., Kanda, K., Kondo, H., Yokoi, T., Kodaira, H., Furuwa, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K., Kamiyama, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T. NEDO human cDNA sequencing project

Unpublished
3 (bases 1 to 2848)
Isogai, T. and Yamamoto, J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FUJ Project (HRI Team); 2-6-7

Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (Fax: 81-438-52-3986) (E-mail: genomics@kai.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES
source

Location/Qualifiers
1. 2848
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="SPLEN2019405"
/tissue type="spleen"
/clone_lib="SPLEN2"
/note="Cloning vector: pME18SFL3"

ORIGIN

Alignment Scores:
Pred. No.: 1.74e-42 Length: 2848
Score: 983.50 Matches: 184
Percent Similarity: 70.1% Conservative: 57
Best Local Similarity: 53.5% Mismatches: 96
Query Match: 34.5% Indels: 7
DB: 8 Gaps: 2

US-10-635-977-2 (1-541) x AK097236 (1-2848)

Qy 54 GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlnLeuHisGlu 73

RESULT 15
BT021865

Db 1294 GAGTTTGTGGATGAAGCTCTGTGTGGTGGAGGAGTACCTTAGCAACTTGGCCACATG 1353
Qy 74 AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrrGluAspLeu 93
Db 1354 GACATCGAACAGGACCTGGAGCCCGCTGTACCTCACCCCGAGGGCTGGTCCCTTTC 1413
Qy 94 ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
Db 1414 CTCACGCGCTACTACCAAGTGTCTCCACGAAGGGGAGAACTCAGGCACCTCCAGACTCAG 1473
Qy 114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
Db 1474 GTCCAGCGCTGTGGAGACATCTCGCAGCAGCTGCAGGCGCTGTACCTCCCGAGATAGACATG 1533
Qy 134 AspGlyLeuArgAsnIleTrrPileLeuLysProAlaAlaLysSerArgGlyArgAspIle 153
Db 1534 GAAGGGATCGCAACATCTGGATCTGAAGCAGGAGCCAGTCCCGTGGAGAGGCAT 1593
Qy 154 ValCysMetAspArgValGluGluLeuLeuAlaAlaAspHisProLeuSer 173
Db 1594 ATGTGCATGGACCACTGGAGGAGATGTGAAGCTGGTGAACGGCAACCCCGTGGTATG 1653
Qy 174 ArgAspAsnLysTrrValValGlnLysTrrIleGluThrProLeuLeuIleCysAspThr 193
Db 1654 AAGGACGGCAAGTGGTGGTGCAGAAAGTATATTGAGCGCGCCCTCTCTATCTTTGGCACC 1713
Qy 194 LysPheAspIleArgGlnTrrPheLeuValThrAspTrrAsnProLeuThrIleTrrPhe 213
Db 1714 AAGTTTGACCTCAGACAGTGGTCTCTGTGTAAGTGAACCCCACTTACCGCTGGTTC 1773
Qy 214 TyrLysGluSerTrrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
Db 1774 TACCGGACAGCTATATCCGCTTTTCCACGCGAGCCCTTCTCCCTGAAGAACCTGGACAC 1833
Qy 234 AlaIleHisLeuCysAsnAlaValGlnLysTrrLeuLysAsnAspValGlyArgSer 253
Db 1834 TCAGTGCACTGTGCAACCACTCCATCCAGAAAGCACTGGAGAACTCATGTCATCGGCAT 1893
Qy 254 ProLeuLeuProAlaHisAsnMetTrrThrSerThrArgPheGlnGluTrrLeuGlnArg 273
Db 1894 CCATGCTTCCGCCAGACCAACATGTGGTCTAGCCAGAGGTTCCAGGCGCCACTCGCAGAG 1953
Qy 274 GlnGlyArgGlyAlaValTrrPglySerValIleTrrProSerMetLysLysAlaIleAla 293
Db 1954 ATGGTGTCCCCAAATGCTTGGTCCACCATCATCTGCTGCATGAAGAGTCTGTGATC 2013
Qy 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
Db 2014 CACGCACTTCAGACCTCCAGGACACCGTGCATGTCCGAAGGCCAGCTTTGAGCTCTAT 2073
Qy 314 GlyAlaAspPheValLeuGlyArgAspPheArgProTrrPheIleGluIleAsnSerSer 333
Db 2074 GGCCTGACTTCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2133
Qy 334 ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353
Db 2134 CCCACGATGGACCCCTCCACAGCAGTCACTGCCCGGCTCTGTGCTGGCTGCAAGCTGAC 2193
Qy 354 ThrIleLysValAlaVal-----AspArgSerCysAspIleGlyAsnPheGlu 369
Db 2194 ACCCTGGCGTGGTCAATTGACCGGAGGCTGGACCGCAACTGTGACAGGAGGCCTTTTGG 2253
Qy 370 LeuLeuTrrArgGln-----ProValValGluProProPheSerGlySerAsp 386
Db 2254 CTCATCTATAGCAGGCGCCAGCTCTCTAAATGCAAGTGTCCACGAGAGAAATGCACCA 2313
Qy 387 LeuCysValAla 390
Db 2314 CTGTGCCAGCA 2325

```

LOCUS      BT021865      1832 bp      mRNA      linear      MAM 17-MAR-2005
DEFINITION Bos taurus tubulin tyrosine ligase-like family, member 3 (TTLL3),
            mRNA, complete cds.
ACCESSION  BT021865
VERSION    BT021865.1  GI:61555425
KEYWORDS   FLI CDNA.
SOURCE     Bos taurus (cow)
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE  1 (bases 1 to 1832)
AUTHORS   Smith,T.P., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
            Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
            Chitko-McKown,C.G., Perte,G., Holt,I., Karamnycheva,S., Liang,F.,
            Quackenbush,J. and Keele,J.W.
TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
JOURNAL   Genome Res. 11 (4), 626-630 (2001)
PUBMED    11282978
REFERENCE  2 (bases 1 to 1832)
AUTHORS   Harhay,G.P., Sonstegard,T.S., Van Tassell,C.P., Clawson,M.L.,
            Heaton,M.P., Keele,J.W., Snelling,W.M., Weidmann,R.T. and
            Smith,T.P.L.
TITLE     Sequencing and analysis of Bos taurus full-length insert cDNA
            clones
JOURNAL   Unpublished
REFERENCE  3 (bases 1 to 1832)
AUTHORS   Harhay,G.P., Sonstegard,T.S., Van Tassell,C.P., Clawson,M.L.,
            Heaton,M.P., Keele,J.W., Snelling,W.M., Weidmann,R.T. and
            Smith,T.P.L.
TITLE     Direct Submission
JOURNAL   Submitted (17-MAR-2005) Molecular Genetic Research Unit, USDA-ARS
            U.S. Meat Animal Research Center, P.O. Box 166, Spur 18D, Clay
            Center, NE 68933
COMMENT   Contact: Gregory P. Harhay at harhay@email.marc.usda.gov.
FEATURES   Location/Qualifiers
            source
            1..1832
            /organism="Bos taurus"
            /mol_type="mRNA"
            /db_xref="taxon:9913"
            /clone="280V1704; plate 17 row O column 4"
            /tissue_type="pooled"
            /lab_host="Escherichia coli DH10B"
            /notes="Vector: pCMV SPOT6; Site 1: NotI; Site 2: SalI;
            Library made from pooled tissue from testis, thymus,
            semitendinosus muscle, longissimus muscle, pancreas,
            adrenal, and endometrium."
            gene
            1..1832
            /gene="TTLL3"
            /notes="putative, similar to GeneID: 26140"
            380..1165
            /gene="TTLL3"
            /codon_start=1
            /product="tubulin tyrosine ligase-like family, member 3"
            /protein_id="AAK46712.1"
            /db_xref="GI:61555426"
            /trnslations="MDMEGRNIIWIKPGAKSRGRCIMCMHLEMLKLVDNPNMVK
            DGKVVHNYIERPLLIFGTFEDLRQWPLTDWNPLOTWVYRDSYIRFSTQPSFLKNLD
            NSVHLNNSIQKLELNSCHRPPLLPDPMWSSQKQAHQETGAPNANSTVIYVGMKA
            AVHALQSDTVQCRKASFELGADVFGEQFPLNIEINASPTWAPSTAVTARLCA
            GVQADTLRVVIDWRLDRITVRGLSSSTSSLPWRCFPMWVSGF"
            ORIGIN
            Alignment Scores:
            Pred. No.:      1.72e-41      Length:      1832
            Score:         960.50         Matches:     224
            Percent Similarity: 53.1%      Conservative: 67
            Best Local Similarity: 40.9%    Mismatches:  169
            Query Match:     33.7%       Indels:      91
            DB:              4           Gaps:        9

```

```

US-10-635-977-2 (1-541) x BT021865 (1-1832)
QY      13  GlnSerCysSerArgSerSerArgSerLysProArgAspGlnArgGluAlaGlySer 32
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      18  AAGTCAAACCTCTGTGCAAGAGAAGAGCGCCCAAGAGACAAGCAG----- 62
QY      33  SerAspLeuSerSerArgGlnAspAlaGluAsnAlaGluAlaLysLeuArgGlyLeuPro 52
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      63  -----CAAAAGAAACAGGAGAAGAACGACGATGTCGGTGTTC 98
QY      53  GlyGlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHis 72
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      99  CCAGAGTTTGTGGATCAGGCTCTGTGCTCGGAGGACACCTTAAACACCTGGCGCAT 158
QY      73  GluAspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAsp 92
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      159  CTGGACATCGACAAGGATGTGGCGACCGCGCTGCTGGGCCCTCGAGGCTGTGTCCTC 218
QY      93  LeuThrGlnGlnTyrTyrSerLeuValHis-----GlyAspAlaPhe----- 106
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      219  TTCCTCCAGCGCTACTACCAAGTGGTFCAGATGCCCGCCAGGCCCGCCCTGCCCAAGTCC 278
QY      107  -----IleSerAsnSerArgAsnTyrPheSerGln 116
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      279  CATCATCTCCCCACAGCGAGGGGGCAGAACTCAGGTACCTCGAC-ACTCAGGTCCAGCGC 337
QY      117  CysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeu 136
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      338  TGTGAGGACATCTCGACGAGCTGCGGGCTGTGTACCCCGACAGATGACATGGAAGAGAT 397
QY      137  ArgAsnIleTrpIleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMet 156
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      398  CCAACATCTGGATCTGAAGCGCGGAGCCAGTCCCGTGGAGAGATGATCATGTGCATG 457
QY      157  AspArgValGluGluLeuLeuAlaAlaAlaAspHisProLeuSerArgAspAsn 176
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      458  GACCACCTGGAGGAATGCTGAAGCTGGTGGAGTGCACCCCATGATGAAGAGCGGC 517
QY      177  LysTrpValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAsp 196
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      518  NAGTGGGTGGTGCAATAATATATATCGAGCGGCCCTCTCATCTTTGGCCACCAAAATTGAT 577
QY      197  IleArgGlnTrpPheIleValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGlu 216
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      578  CTGAGCAGTGGTTCCTGTGTGACTGACTGAAACCCACTTACTGTGTGGTGTCTACCGTGAC 637
QY      217  SerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHis 236
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      638  AGCTACATCGGCTTCTCCACACAGCCCTTCTCCTGAAGAACCTGGACAACTCCGTGCAC 697
QY      237  LeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeu 256
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      698  CTGTGCAACAACATCCATCCAGAAGCACCTGGAGATTTCATGCCACCGGCCACCCACTGCTG 757
QY      257  ProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArg 276
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      758  CCCCCAGACAAACATGGTTCAGCAAGAAATCCAGGCCCTCCAGCCCTCGAGGAGACGGGGGCC 817
QY      277  GlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMet 296
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      818  CCGAACCGCTGGTCCACTGTCTCATCGTCCGTGCGATGAAGGCCCGCAGTGATCCACGCCCTG 877
QY      297  LysValAlaGlnAspHisValGluProArgLysLysAsnSerPheGluLeuTyrGlyAlaAsp 316
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      878  CAGACCTCCAGGACACCGGTGAGTGTGGAAGCCAGCTTCGAGCTCTACGGTGTCTGCTGTC 937
QY      317  PheValLeuGlyArgAspPheArgProTrpLeuLeuGluIleAsnSerSerProThrMet 336
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      938  TTTGTGTTCGGTGAGGACTTCCAGCCCTTGCTGATCGAGATCAACGCCACCCACCATG 997
QY      337  HisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLys 356
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```
Db 998 GCACCTCCACGGCTGTCACCGCAGGCTCTGTGCGGTGTGTGAGGCCGACACCCCTGCGT 1057
Qy 357 ValAlaVal-----AspArgSerCysAspIleGlyAsnPheGluLeuTrp 372
Db 1058 GTGGTCATGACTGGCGACTGGACCGCAC-TGTGACACGGGGCCTTTGAGCTCATCTAC 1116
Qy 373 ArgGlnProValValGluProProPheSerGlySerAspLeuCysValAlaGlyVal 392
Db 1117 AAGCAGCCTGCGGTGAGGTGCCCCCAGTATGTGGGTATCCGGCTTCTAGTAGGGCTCC 1176
Qy 393 SerValArgAlaArgAlaArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSer 412
Db 1177 ACCATCAAGAG-----CCC 1191
Qy 413 LeuLeuAspAlaGlnProLeuLysAlaArgGlyProSerAlaMetProAspProAlaGln 432
Db 1192 CTGGC-GATGAGTCACCGCGGATGCCGTCGCCCCAGC----- 1229
Qy 433 GlyProProSerProAla-LeuGlnArgAspLeuGlyLeuLysGluLysGlyLeuPr 452
Db 1230 ---CCTCCCTCACCTGCTGCCCCGAGAGGCTCTGGGGAAGCAAGGACTCAGGAAGCCC 1286
Qy 452 oLeu----- 453
Db 1287 CATCCACAGCCCCCTTACCACCTTCCCAAGCCTCCACAGCAAGCCCGGCTGCGCTTCTCCCC 1346
Qy 454 -AlaLeuLeuAlaProLeuArgGly-----AlaAlaGluSerGlyGlyAlaAl 469
Db 1347 ATGTGCCCTGGCCCCCAGGGGCAGGTCTCTCAGACTACAGCAGCAGCAAGCTGGTGGGCATA 1406
Qy 469 aGlnProThr-----ArgThrLysAlaAlaGlyLysValGl 481
Db 1407 AGGCCCTGTGACCAACAGAGGCTTGATGACTCTACCAACCGCCAAGGTTTGTATTT 1466
Qy 481 uLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValProVa 501
Db 1467 CCCTCCACCTAACCTGAGCTCAGCTGGCACCACCGCCCTGAACCCAAAAAGGCTC 1526
Qy 501 lAlaGlnProAlaLysSerTrp 508
Db 1527 TTGCTCCCCCAGCGCCTTCATGG 1548
```

Search completed: April 4, 2006, 13:37:14
Job time : 7987 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7

Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:59:01 ; Search time 5.39777 Seconds
(without alignments)
1364.812 Million cell updates/sec

Title: US-10-635-977-2_COPY_133_374

Perfect score: 1299

Sequence: 1 IDGLNIIWPKAKSRGRD.....IKVADRSCDIGNFELLWRQ 242

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 180808 seqs, 30441898 residues

Total number of hits satisfying chosen parameters: 180808

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New:*

- 1: /SID85/prodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /SID85/prodata/2/pubpaa/US06_NEW_PUB.pep.*
- 3: /SID85/prodata/2/pubpaa/US07_NEW_PUB.pep.*
- 4: /SID85/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 5: /SID85/prodata/2/pubpaa/US05_NEW_PUB.pep.*
- 6: /SID85/prodata/2/pubpaa/US10_NEW_PUB.pep.*
- 7: /SID85/prodata/2/pubpaa/US11_NEW_PUB.pep.*
- 8: /SID85/prodata/2/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	271	20.9	439	7	US-11-072-512-3658
2	270	20.8	592	7	US-11-072-512-3371
3	177	13.6	404	7	US-11-072-512-3621
4	79.5	6.1	433	7	US-11-167-836-26
5	79	6.1	462	7	US-11-096-568A-25994
6	79	6.1	470	7	US-11-096-568A-25993
7	79	6.1	517	7	US-11-096-568A-25992
8	79	6.1	549	7	US-11-085-185-2
9	78	6.0	256	7	US-11-167-831-9
10	78	6.0	256	7	US-11-167-831-14
11	78	6.0	667	6	US-10-793-626-198
12	77	5.9	256	5	US-09-810-501-67
13	77	5.9	256	7	US-11-167-831-8
14	77	5.9	256	7	US-11-167-831-11
15	76	5.9	256	7	US-11-167-831-12
16	76	5.9	256	7	US-11-167-831-13
17	75	5.8	362	7	US-11-012-762-56
18	75	5.8	1047	7	US-11-072-512-2408
19	75	5.8	1137	7	US-11-087-099-5799
20	74.5	5.7	646	7	US-11-087-099-10725
21	74.5	5.7	698	7	US-11-087-099-8952
22	74.5	5.7	932	7	US-11-017-550-65
23	74.5	5.7	932	7	US-11-002-387-65
24	74	5.7	179	7	US-11-072-512-2871
25	74	5.7	255	7	US-11-167-831-10

26	73	5.6	351	7	US-11-096-568A-5676	Sequence 5676, Ap
27	73	5.6	436	7	US-11-096-568A-5675	Sequence 5675, Ap
28	73	5.6	497	6	US-10-793-626-2812	Sequence 2812, Ap
29	73	5.6	729	6	US-10-995-561-878	Sequence 878, App
30	73	5.6	744	6	US-10-995-561-876	Sequence 876, App
31	73	5.6	752	6	US-10-995-561-879	Sequence 879, App
32	73	5.6	753	6	US-10-995-561-877	Sequence 877, App
33	73	5.6	1131	7	US-11-087-099-2209	Sequence 2209, Ap
34	73	5.6	1131	7	US-11-087-099-9522	Sequence 9522, Ap
35	72.5	5.6	302	7	US-11-098-686-10927	Sequence 10927, A
36	72.5	5.6	698	7	US-11-087-099-9341	Sequence 9341, Ap
37	72.5	5.6	888	6	US-10-131-826A-544	Sequence 544, App
38	72.5	5.6	888	6	US-10-973-115B-544	Sequence 544, App
39	71.5	5.5	420	7	US-11-096-568A-32310	Sequence 32310, A
40	71.5	5.5	508	7	US-11-096-568A-32309	Sequence 32309, A
41	71.5	5.5	508	7	US-11-172-740-239	Sequence 239, App
42	71.5	5.5	508	7	US-11-172-740-300	Sequence 300, App
43	71.5	5.5	512	7	US-11-096-568A-32308	Sequence 32308, A
44	71	5.5	471	7	US-11-024-959-399	Sequence 399, App
45	70.5	5.4	121	7	US-11-107-028-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1

US-11-072-512-3658
; Sequence 3658, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3658
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-3658

Query Match 20.9%; Score 271; DB 7; Length 439;

Best Local Similarity 29.8%; Pred. No. 6.3e-20;

Matches 75; Conservative 45; Mismatches 86; Indels 46; Gaps 9;

Qy 8 WIIPAKSRGRDVCMDRVEEILEAAADHPLSRDNK-----WVQKVIETPLICD 60

Db 164 WIMPKVARSOKGLFFRLKLDIVDM-RKOTRSSDDQDDIPVNYVAQRVIENPYLIGG 222

Qy 61 TKFDIRQWFLVTWNPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCNNAVOKLNDVGR 120


```
QY 60 DTQFDIRQWFLVTDWNL-----TIWFKESYLK-----PSTQRFSLDKLDSAIHLCCNAV 110
Db 93 DAAFSRPRFAVGKYTAYDCSGLWAPFEPYLRQARRICATELFSATRLUESFHIIRDEEV 152
QY 111 Q---KYLKNDVGRSPLLPAAHNMWTSRFOEYLQGRGAVGWGVIYSPMKKAIAHAMKVA 167
Db 153 RVMRLQRQAAGR-----TVRLRDLQMLALGVISRVL---GKK---YVMEEA 195
QY 168 QD-----HVEPRK-----NSFELYGADFVLGRDRPWL 195
Db 196 ADGEDSAPAITPAEFREMWDEFFALHGA-FNIG-DVIPWL 234

RESULT 8
US-11-085-185-2
; Sequence 2, Application US/11085185
; Publication No. US20050287634A1
; GENERAL INFORMATION:
; APPLICANT: MACHIDA, Masayuki
; TITLE OF INVENTION: NOVEL GENES OF CELL WALL-DEGRADING ENZYMES DERIVED FROM ASPERGILL
; TITLE OF INVENTION: METHOD FOR THE PRODUCTION OF THE ENZYMES
; FILE REFERENCE: 4600-0110PUS1
; CURRENT APPLICATION NUMBER: US/11/085,185
; CURRENT FILING DATE: 2005-03-22
; PRIOR APPLICATION NUMBER: JP 2004-188849
; PRIOR FILING DATE: 2004-06-25
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 549
; TYPE: PRT
; ORGANISM: Aspergillus oryzae
US-11-085-185-2

Query Match 6.1%; Score 79; DB 7; Length 549;
Best Local Similarity 21.2%; Pred. No. 4.4; Indels 34; Gaps 6;
Matches 38; Conservative 28; Mismatches 79;

QY 27 VEEILELAAADHPLSRDNKQWVQKY-IETPLLCIDTKFDIRQWFLVTDWNLPTIWFYKES 85
Db 323 VEDVLN-----WTTDATKLQKNGISAVYTVHDGFLNLSKWQMLKDRPDRMFLDTHQ 375
QY 86 YLRSTQRFSLDKLDSAIHLCN---NAVQYKLVNDVGRSPLLPAAHNMWTSRFOEYLQGR 142
Db 376 YTIENTCQIVLNHPTDRVKLICNDWYNNMIKEINTTSAGWGPTICGWSQADTDCAQYLLNV 435
QY 143 GRGAVW-----GSVIY-----PSMKKAIAHA-----MKVAQDHVEPRKNSF 178
Db 436 GRGTRWEGTFAIGDSTVCPTADTGPTCSCASANAPPADYSDDGKYKFLQTYAEAQMSAF 494

RESULT 9
US-11-167-831-9
; Sequence 9, Application US/11167831
; Publication No. US20060062805A1
; GENERAL INFORMATION:
; APPLICANT: PAUL DR, PREM S
; APPLICANT: ZHANG, YANJIN
; TITLE OF INVENTION: PROTEINS ENCODED BY POLYNUCLEIC ACIDS OF PORCINE
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PPRSV)
; FILE REFERENCE: 8199-0005-55XCIP WO
; CURRENT APPLICATION NUMBER: US/11/167,831
; CURRENT FILING DATE: 2005-06-28
; PRIOR APPLICATION NUMBER: US/10/428,826
; PRIOR FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: PCT/US99/02630
; PRIOR FILING DATE: 1999-04-19
; PRIOR APPLICATION NUMBER: US 09/019,793
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: US 08/478,316
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/301,435
; PRIOR FILING DATE: 1994-09-01
; PRIOR APPLICATION NUMBER: US 08/131,625
; PRIOR FILING DATE: 1993-10-05
; PRIOR APPLICATION NUMBER: US 07/969,071
; PRIOR FILING DATE: 1992-10-30
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Porcine reproductive and respiratory syndrome virus
```

```
; PRIOR APPLICATION NUMBER: US 08/301,435
; PRIOR FILING DATE: 1994-09-01
; PRIOR APPLICATION NUMBER: US 08/131,625
; PRIOR FILING DATE: 1993-10-05
; PRIOR APPLICATION NUMBER: US 07/969,071
; PRIOR FILING DATE: 1992-10-30
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Porcine reproductive and respiratory syndrome virus
US-11-167-831-9

Query Match 6.0%; Score 78; DB 7; Length 256;
Best Local Similarity 18.8%; Pred. No. 2.1;
Matches 48; Conservative 34; Mismatches 73; Indels 100; Gaps 12;

QY 46 WVVKYIETPLLI-----CDTKFDIRQWFLVTDW-----NPLTIWFKESYLK 88
Db 18 WMLSRSSWCPLLISLYFWFPCLASPSPVGWGWSFASDFAPRYSVRALPFTLSNYSRYSEA 77
QY 89 FSTQRFSLDKLDSAIHLCCNAVQKY-LKNDVGRSPLLPAAHNMWTSRFOEYLQGR 141
Db 78 FLSQ-----CQVDIPTWGTGKHPG---MLWHKVVSTLIDEMVSRMYRIMEK 121
QY 142 QGRGAVGWGVIYP-----SMKKAIAHAMKVAQDHVE-----PRKNSFELYGADFV 186
Db 122 AGQ-AAWKQVSEATLSRISLSDVAHFQHLAAIEAEATCKYLASRLPMLHNLRTGNSVT 180
QY 187 LG-----RDFRPWLIEINSSPTMHPSTPVTVAQLCAQVQEDTIKV 225
Db 181 IVYNSTLNQVFAIFPTGSRPKLHDFQOQLIAVHSS-----IFS 219
QY 226 AVDRSCDIGNFELLW 240
Db 220 SVAASCTL--FVVLW 232

RESULT 10
US-11-167-831-14
; Sequence 14, Application US/11167831
; Publication No. US20060062805A1
; GENERAL INFORMATION:
; APPLICANT: PAUL DR, PREM S
; APPLICANT: ZHANG, YANJIN
; TITLE OF INVENTION: PROTEINS ENCODED BY POLYNUCLEIC ACIDS OF PORCINE
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PPRSV)
; FILE REFERENCE: 8199-0005-55XCIP WO
; CURRENT APPLICATION NUMBER: US/11/167,831
; CURRENT FILING DATE: 2005-06-28
; PRIOR APPLICATION NUMBER: US/10/428,826
; PRIOR FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: US/09/601,326
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/02630
; PRIOR FILING DATE: 1999-04-19
; PRIOR APPLICATION NUMBER: US 09/019,793
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: US 08/478,316
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/301,435
; PRIOR FILING DATE: 1994-09-01
; PRIOR APPLICATION NUMBER: US 08/131,625
; PRIOR FILING DATE: 1993-10-05
; PRIOR APPLICATION NUMBER: US 07/969,071
; PRIOR FILING DATE: 1992-10-30
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Porcine reproductive and respiratory syndrome virus
```


US-11-167-831-14

Query Match 6.0%; Score 78; DB 7; Length 256;
Best Local Similarity 18.8%; Pred. No. 2.1;
Matches 48; Conservative 34; Mismatches 73; Indels 100; Gaps 12;

QY 46 WVQKYEITPLLI-----CDTKFDIRQWFLVTDW-----NPLTIWFYKESYLR 88
DB 18 WMLSRSSWCPLLISLYFWPFCCLASPSQVGMWSPASDFAPRYSVRALPFTLSNRYRSEYA 77
QY 89 FSTQRFSLDKLDSAIHLCCNNAVQKY-LKNDVGRSPLLPAHNMWT-----STRFQYLOR 141
DB 78 FLSQ-----CQVDIPTWGTGKPLG---MLMHHKVTSLIDEMVSRMYRIMEK 121
QY 142 QGRGAVGMSVIVP-----SMKKAIAHAMKVAQDHVE-----PRKNSFELYGADFV 186
DB 122 AGQ-AAWKQVSEATLSRISLSDVVAHFQHLAAIEAETCKYLASRLPMLHLNLRMTGSNVT 180
QY 187 LG-----RDRPWLIEINSSPTMHPSTPVTVAQLCAQVQEDTIKV 225
DB 181 IVYNSTLNQVFAIFPTPGSRPKLHDFQWLIAVHSS-----IFS 219
QY 226 AVDRSCDIGNFELLW 240
DB 220 SVAASCTL--FVVLW 232

RESULT 11

US-10-793-626-198
; Sequence 198, Application US/10793626
; Publication No. US20050255478A1

; GENERAL INFORMATION:

; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 198
; LENGTH: 667

; TYPE: PRT
; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence

US-10-793-626-198

Query Match 6.0%; Score 78; DB 6; Length 667;
Best Local Similarity 24.0%; Pred. No. 7.2;
Matches 41; Conservative 34; Mismatches 70; Indels 26; Gaps 9;

QY 73 DWN-----PLTIWFYKESYLRSTQRFSLDKLDSAIHLCCNNAVQKYLRKNDVGRSPLLPAH 128
DB 15 EWNLDELPIIPKQMKTVISKTNRQF---KIVKSLIDKNVKEIIATDAGREGELVARL 71
QY 129 MWTSTRFQYLORQGRGAVGMSVIVPSSMKKAIAHAMKVAQDHVEPRKNSFE-LYCADFVL 187
DB 72 ILDKVGNKKPIKR-----LWISSV---TKKAIQEGFKLKN-----GNAYQNLVEA--AL 116
QY 188 GRDRPWLIEINSSPTMHPSTPVTVAQL-CAQVQEDTIKVAVDRSCDIGNFE 237
DB 117 ARSEADWIVGINATFAL--TTKYDAQLSLGRVQPTTQIVKSRQDEINFK 165

RESULT 12

US-09-810-501-67

; Sequence 67, Application US/09810501
; Publication No. US20060029928A1

; GENERAL INFORMATION:

; APPLICANT: PAUL, PREM S.

MENG, XIANG-JIN
HALBUR, PATRICK G.
MOROZOV, IGOR
LUM, MELISSA A.

TITLE OF INVENTION: A POLYNUCLEIC ACID ISOLATED FROM A

PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRS)
A PROTEIN ENCODED BY THE POLYNUCLEIC ACID, A VACCINE
PREPARED FROM OR CONTAINING THE POLYNUCLEIC ACID OR

NUMBER OF SEQUENCES: 77

CORRESPONDENCE ADDRESS:

ADDRESSER: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

P.C.

STREET: 1755 S. Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/810,501

FILING DATE: 12-Mar-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/301,435

FILING DATE: <Unknown>

APPLICATION NUMBER: 08/131,625

FILING DATE: 05-OCT-1993

ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye, Jean-Paul M.P.

REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 4625-021-55X CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 67:

SEQUENCE CHARACTERISTICS:

LENGTH: 256 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 67:

US-09-810-501-67

Query Match 5.9%; Score 77; DB 5; Length 256;
Best Local Similarity 18.8%; Pred. No. 2.7;
Matches 48; Conservative 33; Mismatches 74; Indels 100; Gaps 12;

QY 46 WVQKYEITPLLI-----CDTKFDIRQWFLVTDW-----NPLTIWFYKESYLR 88
DB 18 WMLSRSSWCPLLISLYFWPFCCLASPSQVGMWSPASDFAPRYSVRALPFTLSNRYRSEYA 77
QY 89 FSTQRFSLDKLDSAIHLCCNNAVQKY-LKNDVGRSPLLPAHNMWT-----STRFQYLOR 141
DB 78 FLSQ-----CQVDIPTWGTGKPLG---MLMHHKVTSLIDEMVSRMYRIMEK 121
QY 142 QGRGAVGMSVIVP-----SMKKAIAHAMKVAQDHVE-----PRKNSFELYGADFV 186
DB 122 AGQ-AAWKQVSEATLSRISLSDVVAHFQHLAAIEAETCKYLASRLPMLHLNLRMTGSNVT 180
QY 187 LG-----RDRPWLIEINSSPTMHPSTPVTVAQLCAQVQEDTIKV 225
DB 181 IVYNSTLNQVFAIFPTPGSRPKLHDFQWLIAVHSS-----IFS 219
QY 226 AVDRSCDIGNFELLW 240
DB 220 SVAASCTL--FVVLW 232

```
RESULT 13
US-11-167-831-8
; Sequence 8, Application US/11167831
; Publication No. US20060062805A1
; GENERAL INFORMATION:
; APPLICANT: PAUL DR, PREM S
; APPLICANT: ZHANG, YANJIN
; TITLE OF INVENTION: PROTEINS ENCODED BY POLYNUCLEIC ACIDS OF PORCINE
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PPRSV)
; FILE REFERENCE: 8199-0005-55XCIP WO
; CURRENT APPLICATION NUMBER: US/11/167,831
; CURRENT FILING DATE: 2005-06-28
; PRIOR APPLICATION NUMBER: US/10/428,826
; PRIOR FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: US/09/601,326
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/02630
; PRIOR FILING DATE: 1999-04-19
; PRIOR APPLICATION NUMBER: US 09/019,793
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: US 08/478,316
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/301,435
; PRIOR FILING DATE: 1994-09-01
; PRIOR APPLICATION NUMBER: US 08/131,625
; PRIOR FILING DATE: 1993-10-05
; PRIOR APPLICATION NUMBER: US 07/969,071
; PRIOR FILING DATE: 1992-10-30
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Porcine reproductive and respiratory syndrome virus
US-11-167-831-8

Query Match 5.9%; Score 77; DB 7; Length 256;
Best Local Similarity 18.8%; Pred. No. 2.7; Indels 100; Gaps 12;
Matches 48; Conservative 33; Mismatches 74;

QY 46 WVQKYIETPLLI-----CDTKFDIQWFLVTDW-----NPLTIWFKESYL 88
Db 18 WMLSRSSWCPLLSLYFWPFCCLASPSQVGWGFASDFAPRYSVRALPFTLSNYSRYEA 77
QY 89 FSTQRFSLDKLDSAIHCNNNAVQKY-LKNDVGSRPLLPAINMT-----STRFQEV 141
Db 78 FLNQ-----CQVDIPTWGTGKPLG--MLWHHKVSTLIDEMVSRMYRIMEK 121
QY 142 QGRGAVGWSVIYP-----SMKKAIAHAMKVAQDHVE-----PRKNSFELYGADFV 186
Db 122 AGQ-AAWKQVSEATLSRISLSDVVAHFQHLAAIEAETCKYLASRLPMLHMLRMTGSNVT 180
QY 187 LG-----RDFRPMLEINSPTMHPSTPTVTAQLCAQVQEDTIKV 225
Db 181 IVYNSTLNQVFAVPTPGSRPKLHDFQWLIHVS-----IFS 219
QY 226 AVDRSCDIGNFELW 240
Db 220 SVAASCTL--FVVLW 232

RESULT 14
US-11-167-831-11
; Sequence 11, Application US/11167831
; Publication No. US20060062805A1
; GENERAL INFORMATION:
; APPLICANT: PAUL DR, PREM S
; APPLICANT: ZHANG, YANJIN
; TITLE OF INVENTION: PROTEINS ENCODED BY POLYNUCLEIC ACIDS OF PORCINE
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PPRSV)
; FILE REFERENCE: 8199-0005-55XCIP WO
; CURRENT APPLICATION NUMBER: US/11/167,831
; CURRENT FILING DATE: 2005-06-28
```

```
; PRIOR APPLICATION NUMBER: US/10/428,826
; PRIOR FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: US/09/601,326
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/02630
; PRIOR FILING DATE: 1999-04-19
; PRIOR APPLICATION NUMBER: US 09/019,793
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: US 08/478,316
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/301,435
; PRIOR FILING DATE: 1994-09-01
; PRIOR APPLICATION NUMBER: US 08/131,625
; PRIOR FILING DATE: 1993-10-05
; PRIOR APPLICATION NUMBER: US 07/969,071
; PRIOR FILING DATE: 1992-10-30
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Porcine reproductive and respiratory syndrome virus
US-11-167-831-11

Query Match 5.9%; Score 77; DB 7; Length 256;
Best Local Similarity 18.4%; Pred. No. 2.7; Indels 100; Gaps 12;
Matches 47; Conservative 35; Mismatches 73;

QY 46 WVQKYIETPLLI-----CDTKFDIQWFLVTDW-----NPLTIWFKESYL 88
Db 18 WMLSRSSWCPLLSLYFWPFCCLASPSQVGWGFASDFAPRYSVRALPFTLSNYSRYEA 77
QY 89 FSTQRFSLDKLDSAIHCNNNAVQKY-LKNDVGSRPLLPAINMT-----STRFQEV 141
Db 78 FLNQ-----CQVDIPTWGTGKPLG--MFWHHKVSTLIDEMVSRMYRIMEK 121
QY 142 QGRGAVGWSVIYP-----SMKKAIAHAMKVAQDHVE-----PRKNSFELYGADFV 186
Db 122 AGQ-AAWKQVSEATLSRISLSDVVAHFQHLAAIEAETCKYLASRLPMLHMLRMTGSNVT 180
QY 187 LG-----RDFRPMLEINSPTMHPSTPTVTAQLCAQVQEDTIKV 225
Db 181 IVYNSTLNQVLAIFPTGSRPKLHDFQWLIHVS-----IFS 219
QY 226 AVDRSCDIGNFELW 240
Db 220 SVAASCTL--FVVLW 232

RESULT 15
US-11-167-831-12
; Sequence 12, Application US/11167831
; Publication No. US20060062805A1
; GENERAL INFORMATION:
; APPLICANT: PAUL DR, PREM S
; APPLICANT: ZHANG, YANJIN
; TITLE OF INVENTION: PROTEINS ENCODED BY POLYNUCLEIC ACIDS OF PORCINE
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PPRSV)
; FILE REFERENCE: 8199-0005-55XCIP WO
; CURRENT APPLICATION NUMBER: US/11/167,831
; CURRENT FILING DATE: 2005-06-28
; PRIOR APPLICATION NUMBER: US/10/428,826
; PRIOR FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: US/09/601,326
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/02630
; PRIOR FILING DATE: 1999-04-19
; PRIOR APPLICATION NUMBER: US 09/019,793
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: US 08/478,316
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/301,435
; PRIOR FILING DATE: 1994-09-01
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:55:16 ; Search time 34.8606 Seconds
(without alignments)
2900.544 Million cell updates/sec

Title: US-10-635-977-2_COPY_133_374
Perfect score: 1299
Sequence: 1 IDGLNIIWIKPAKSRGRD.....IKVAVDRSCDIGNFELLWRQ 242

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA Main:
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pap.*
2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pap.*
3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pap.*
4: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pap.*
5: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pap.*
6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1299	100.0	242	4	US-10-615-659-4
2	1299	100.0	242	4	US-10-635-977-4
3	1299	100.0	541	4	US-10-615-659-2
4	1299	100.0	541	4	US-10-615-659-13
5	1299	100.0	541	4	US-10-635-977-2
6	1299	100.0	541	4	US-10-635-977-13
7	1247	96.0	293	4	US-10-615-659-14
8	1247	96.0	293	4	US-10-635-977-14
9	850	65.4	352	4	US-10-615-659-7
10	850	65.4	352	4	US-10-635-977-7
11	850	65.4	352	5	US-10-756-149-5267
12	850	65.4	362	3	US-09-925-298-617
13	850	65.4	362	4	US-10-102-806-617
14	850	65.4	488	4	US-10-275-595A-5
15	847	65.4	326	4	US-10-108-260A-4209
16	846	65.1	292	4	US-10-615-659-5
17	846	65.1	292	4	US-10-635-977-5
18	628.5	48.4	992	6	US-11-097-143-23727
19	499.5	38.5	719	6	US-11-097-143-23415
20	337	25.9	566	5	US-10-450-763-36330
21	324	24.9	496	6	US-11-097-143-19014
22	319	24.6	330	4	US-11-0615-659-26
23	319	24.6	330	4	US-10-635-977-26
24	301	23.2	268	4	US-10-424-599-205823
25	297.5	22.9	827	6	US-11-097-143-9312
26	296	22.8	1226	5	US-10-756-149-5184
27	295.5	22.7	423	4	US-10-615-659-8

28	295.5	22.7	423	4	US-10-635-977-8	Sequence 8, Appli
29	284.5	21.9	989	6	US-11-097-143-7680	Sequence 7680, Ap
30	276	21.2	487	6	US-11-097-143-4080	Sequence 4080, Ap
31	274	21.1	49	4	US-10-615-659-21	Sequence 21, Appl
32	274	21.1	49	4	US-10-615-659-22	Sequence 22, Appl
33	274	21.1	49	4	US-10-635-977-21	Sequence 21, Appl
34	274	21.1	49	4	US-10-635-977-22	Sequence 22, Appl
35	271	20.9	439	4	US-10-104-047-3658	Sequence 3658, Ap
36	270	20.8	524	3	US-09-864-761-38213	Sequence 38213, A
37	270	20.8	592	4	US-10-104-047-3371	Sequence 3371, Ap
38	270	20.8	917	6	US-11-097-143-21303	Sequence 21303, A
39	258	19.9	92	4	US-10-424-599-262294	Sequence 262294, A
40	247.5	19.1	379	4	US-10-615-659-6	Sequence 6, Appli
41	247.5	19.1	379	4	US-10-635-977-6	Sequence 118, App
42	244.5	18.8	377	4	US-10-210-130-118	Sequence 1, Appli
43	244.5	18.8	377	4	US-10-250-613-1	Sequence 2495, Ap
44	205.5	15.8	553	4	US-10-108-260A-2495	Sequence 20, Appl
45	205.5	15.8	887	4	US-10-473-574-20	

ALIGNMENTS

RESULT 1
US-10-615-659-4
; Sequence 4, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-4

Query Match		100.0%;	Score 1299;	DB 4;	Length 242;
Best Local Similarity		100.0%;	Pred. No. 3.2e-129;		
Matches 242;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	IDGLNIIWIKPAKSRGRDIVCM	DRVBEILELAADHPLSRDNK	WVQKYIETPLLICD	60
DB	1	IDGLNIIWIKPAKSRGRDIVCM	DRVBEILELAADHPLSRDNK	WVQKYIETPLLICD	60
QY	61	TKFDIRQWFLVTDNPLTIWFY	KESYLRSTORFSLDKLDS	AIHLNNAVKYKNDVGR	120
DB	61	TKFDIRQWFLVTDNPLTIWFY	KESYLRSTORFSLDKLDS	AIHLNNAVKYKNDVGR	120
QY	121	SPLLPAHNNWVSTRFOEYLQ	RGVAVGVSIVYPSMKKAIA	HAMKVAQDHVPRKNSFEL	180
DB	121	SPLLPAHNNWVSTRFOEYLQ	RGVAVGVSIVYPSMKKAIA	HAMKVAQDHVPRKNSFEL	180
QY	181	YGADPVLGRDPRPWLIEINS	SPTMHPSTPVTVAQLCAQ	QVQEDTIKVAVDRSCDIGN	240
DB	181	YGADPVLGRDPRPWLIEINS	SPTMHPSTPVTVAQLCAQ	QVQEDTIKVAVDRSCDIGN	240
QY	241	RQ	242		
DB	241	RQ	242		

RESULT 2
US-10-635-977-4
; Sequence 4, Application US/10635977
; Publication No. US20040171131A1

GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-4

Query Match 100.0%; Score 1299; DB 4; Length 242;
Best Local Similarity 100.0%; Pred. No. 3.2e-129;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IDGLRNIIWKPAKSRGRDIVCMDRVVEEILELAADHPILSRDNKVVQKIETPLICD 60
DB 1 IDGLRNIIWKPAKSRGRDIVCMDRVVEEILELAADHPILSRDNKVVQKIETPLICD 60
QY 61 TKFDIRQWFLVTDWNPITWIFKESYLRFTSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 120
DB 61 TKFDIRQWFLVTDWNPITWIFKESYLRFTSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 120
QY 121 SPLLPAHNMTSTRFQYLRQGRGAVWGSVIYPSMKKAIHAHMKVAQDHVEPRKNSFEL 180
DB 121 SPLLPAHNMTSTRFQYLRQGRGAVWGSVIYPSMKKAIHAHMKVAQDHVEPRKNSFEL 180
QY 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDKSCDIGNFELLW 240
DB 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDKSCDIGNFELLW 240
QY 241 RQ 242
DB 241 RQ 242

RESULT 3
US-10-615-659-2
; Sequence 2, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-2

Query Match 100.0%; Score 1299; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 9.5e-129;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IDGLRNIIWKPAKSRGRDIVCMDRVVEEILELAADHPILSRDNKVVQKIETPLICD 60
DB 133 IDGLRNIIWKPAKSRGRDIVCMDRVVEEILELAADHPILSRDNKVVQKIETPLICD 192

QY 61 TKFDIRQWFLVTDWNPITWIFKESYLRFTSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 120
DB 193 TKFDIRQWFLVTDWNPITWIFKESYLRFTSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 252
QY 121 SPLLPAHNMTSTRFQYLRQGRGAVWGSVIYPSMKKAIHAHMKVAQDHVEPRKNSFEL 180
DB 253 SPLLPAHNMTSTRFQYLRQGRGAVWGSVIYPSMKKAIHAHMKVAQDHVEPRKNSFEL 312
QY 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDKSCDIGNFELLW 240
DB 313 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDKSCDIGNFELLW 372
QY 241 RQ 242
DB 373 RQ 374

RESULT 4
US-10-615-659-13
; Sequence 13, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-13

Query Match 100.0%; Score 1299; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 9.5e-129;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IDGLRNIIWKPAKSRGRDIVCMDRVVEEILELAADHPILSRDNKVVQKIETPLICD 60
DB 133 IDGLRNIIWKPAKSRGRDIVCMDRVVEEILELAADHPILSRDNKVVQKIETPLICD 192
QY 61 TKFDIRQWFLVTDWNPITWIFKESYLRFTSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 120
DB 193 TKFDIRQWFLVTDWNPITWIFKESYLRFTSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 252
QY 121 SPLLPAHNMTSTRFQYLRQGRGAVWGSVIYPSMKKAIHAHMKVAQDHVEPRKNSFEL 180
DB 253 SPLLPAHNMTSTRFQYLRQGRGAVWGSVIYPSMKKAIHAHMKVAQDHVEPRKNSFEL 312
QY 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDKSCDIGNFELLW 240
DB 313 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDKSCDIGNFELLW 372
QY 241 RQ 242
DB 373 RQ 374

RESULT 5
US-10-635-977-2
; Sequence 2, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977

; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-2

Query Match 100.0%; Score 1299; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 9.5e-129;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IDGLRNIIWIKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKYEITPLICD 60
DB 133 IDGLRNIIWIKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKYEITPLICD 192
QY 61 TKFDIRQWFLVTDNPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 120
DB 193 TKFDIRQWFLVTDNPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 252
QY 121 SP LLPAHNNMTSTRFOEYLQRCRGAVGWSVIYPSMKKAIAMKVAQDHVEPRKNSFEL 180
DB 253 SP LLPAHNNMTSTRFOEYLQRCRGAVGWSVIYPSMKKAIAMKVAQDHVEPRKNSFEL 312
QY 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVD RSCDIGNFELLW 240
DB 313 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVD RSCDIGNFELLW 372
QY 241 RQ 242
DB 373 RQ 374

RESULT 6
US-10-635-977-13
; Sequence 13, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-13

Query Match 100.0%; Score 1299; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 9.5e-129;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IDGLRNIIWIKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKYEITPLICD 60
DB 133 IDGLRNIIWIKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKYEITPLICD 192
QY 61 TKFDIRQWFLVTDNPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 120
DB 193 TKFDIRQWFLVTDNPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 252

QY 121 SP LLPAHNNMTSTRFOEYLQRCRGAVGWSVIYPSMKKAIAMKVAQDHVEPRKNSFEL 180
DB 253 SP LLPAHNNMTSTRFOEYLQRCRGAVGWSVIYPSMKKAIAMKVAQDHVEPRKNSFEL 312
QY 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVD RSCDIGNFELLW 240
DB 313 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVD RSCDIGNFELLW 372
QY 241 RQ 242
DB 373 RQ 374

RESULT 7
US-10-615-659-14
; Sequence 14, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-14

Query Match 96.0%; Score 1247; DB 4; Length 293;
Best Local Similarity 100.0%; Pred. No. 1.4e-123;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IDGLRNIIWIKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKYEITPLICD 60
DB 61 IDGLRNIIWIKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKYEITPLICD 120
QY 61 TKFDIRQWFLVTDNPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 120
DB 121 TKFDIRQWFLVTDNPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCCNNAVQKYLKNDVGR 180
QY 121 SP LLPAHNNMTSTRFOEYLQRCRGAVGWSVIYPSMKKAIAMKVAQDHVEPRKNSFEL 180
DB 181 SP LLPAHNNMTSTRFOEYLQRCRGAVGWSVIYPSMKKAIAMKVAQDHVEPRKNSFEL 240
QY 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVD RSCDI 233
DB 241 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVD RSCDI 293

RESULT 8
US-10-635-977-14
; Sequence 14, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14

Db 61 MEGDRNIWIKPAKSRGRGIMCNDHLEMLKLVNGNPPVWMDKGVVQKVIETPLIFG 120
QY 61 TKFDIROMFLVTDNPLTIWFKESYLRFSTQSFSLDKLDSAIHLCCNAVQKYLKNDVGR 120
Db 121 TKFDLRQWFLVTDNPLTIWFKESYLRFSTQSFSLDKLDSAIHLCCNAVQKYLKNDVGR 180
QY 121 SPLLPAHNMWTSRFEYLQROGCGAVMGSVIYPSMKKATAHAKVQAODHVEPRKNSFEL 180
Db 181 HPLPPDNWSSORFOHLEMLKLVNGNPPVWMDKGVVQKVIETPLIFG 240
QY 181 YGADFVLGRDPRFWLIEINSSPTMHPSTPVTALCAQVQEDTIKAVV----DRSCDIGNF 236
Db 241 YGADVFVGEDFQFWLIEINASSPTMAPSTAVTARLCAGVQADTLRVVIDRLDRNCDTGAF 300
QY 237 ELLWRQ 242
Db 301 ELIYKQ 306

RESULT 12

US-09-925-298-617
; Sequence 617, Application US/09925298
; Publication No. US20020039764A1
; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: PA103

; CURRENT APPLICATION NUMBER: US/09/925,298

; CURRENT FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05881

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

; NUMBER OF SEQ ID NOS: 846

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 617

; LENGTH: 362

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (307)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-298-617

Query Match 65.4%; Score 850; DB 3; Length 362;
Best Local Similarity 61.8%; Pred. No. 2.7e-81;
Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1;

QY 1 IDGLRNITIKPAKSRGRDIVCMRDRVEILELAADHPLSRDNKVVQKVIETPLIFG 60
Db 78 MEGDRNIWIKPAKSRGRGIMCNDHLEMLKLVNGNPPVWMDKGVVQKVIETPLIFG 137
QY 61 TKFDIROMFLVTDNPLTIWFKESYLRFSTQSFSLDKLDSAIHLCCNAVQKYLKNDVGR 120
Db 138 TKFDLRQWFLVTDNPLTIWFKESYLRFSTQSFSLDKLDSAIHLCCNAVQKYLKNDVGR 197
QY 121 SPLLPAHNMWTSRFEYLQROGCGAVMGSVIYPSMKKATAHAKVQAODHVEPRKNSFEL 180
Db 198 HPLPPDNWSSORFOHLEMLKLVNGNPPVWMDKGVVQKVIETPLIFG 257
QY 181 YGADFVLGRDPRFWLIEINSSPTMHPSTPVTALCAQVQEDTIKAVV----DRSCDIGNF 236
Db 258 YGADVFVGEDFQFWLIEINASSPTMAPSTAVTARLCAGVQADTLRVVIDRLDRNCDTGAF 317
QY 237 ELLWRQ 242
Db 318 ELIYKQ 323

RESULT 13

US-10-102-806-617
; Sequence 617, Application US/10102806

; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103P1C1
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 617
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (307)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-102-806-617

Query Match 65.4%; Score 850; DB 4; Length 362;
Best Local Similarity 61.8%; Pred. No. 2.7e-81;
Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1;
QY 1 IDGLRNITIKPAKSRGRDIVCMRDRVEILELAADHPLSRDNKVVQKVIETPLIFG 60
Db 78 MEGDRNIWIKPAKSRGRGIMCNDHLEMLKLVNGNPPVWMDKGVVQKVIETPLIFG 137
QY 61 TKFDIROMFLVTDNPLTIWFKESYLRFSTQSFSLDKLDSAIHLCCNAVQKYLKNDVGR 120
Db 138 TKFDLRQWFLVTDNPLTIWFKESYLRFSTQSFSLDKLDSAIHLCCNAVQKYLKNDVGR 197
QY 121 SPLLPAHNMWTSRFEYLQROGCGAVMGSVIYPSMKKATAHAKVQAODHVEPRKNSFEL 180
Db 198 HPLPPDNWSSORFOHLEMLKLVNGNPPVWMDKGVVQKVIETPLIFG 257
QY 181 YGADFVLGRDPRFWLIEINSSPTMHPSTPVTALCAQVQEDTIKAVV----DRSCDIGNF 236
Db 258 YGADVFVGEDFQFWLIEINASSPTMAPSTAVTARLCAGVQADTLRVVIDRLDRNCDTGAF 317
QY 237 ELLWRQ 242
Db 318 ELIYKQ 323

RESULT 14

US-10-275-595A-5
; Sequence 5, Application US/10275595A
; Publication No. US20040078804A1
; GENERAL INFORMATION:

; APPLICANT: YUE, Henry

; APPLICANT: TANG, Y. Tom

; APPLICANT: AU-YOUNG, Janice

; APPLICANT: LU, Dying Aina M.

; APPLICANT: BAUGHN, Mariah R.

; APPLICANT: HILLMAN, Jennifer L.

; APPLICANT: AZIMZAI, Valda

; APPLICANT: LAL, Preeti

; APPLICANT: YAO, Monique G.

; APPLICANT: BANDMAN, Olga

; APPLICANT: BURFORD, Neil

; APPLICANT: BATRA, Sajeev

; APPLICANT: KEARNEY, Liam

; APPLICANT: POLICKY, Jennifer L.

; TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS

; FILE REFERENCE: PF-0772 USN

; CURRENT APPLICATION NUMBER: US/10/275,595A

; CURRENT FILING DATE: 2003-06-13

```
; PRIOR APPLICATION NUMBER: US 60/201,960
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/202,729
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 60/209,705
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 60/210,149
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: US 60/213,215
; PRIOR FILING DATE: 2000-06-21
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 2156553CD1
; US-10-275-595A-5

Query Match          65.4%; Score 850; DB 4; Length 488;
Best Local Similarity 61.8%; Pred. No. 4.1e-81;
Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1;

QY 1 IDGRLNIWIIPKAAGRGDIVCMRVVEEILEAAADHPLSRDNKVVQKYIETPLLCD 60
DB 61 MEGDRNIWIKPGAKSRGRGIMCWDHLEMLKLVNGNPVVMKDGKVVQKYIERPLIFG 120
QY 61 TKFDIROWFLVTDWNLTIWFYKESYLRFTSTQRFSLKNDNSVHLCNNAVQKYLKNDVGR 120
DB 121 TKFDLRQWFLVTDWNLTIWVYRDSYIRFSTQPFSLKNDNSVHLCNNSIQKLENSCHR 180
QY 121 SPLLPAHNNMTSTRFOEYLQROGRGAVWGSVIYPSMKAIAHAMKVAQDHVEPRKNSFEL 180
DB 181 HPLLPPDNMWSQRFQAHLOEMGAPNAWSTIIIVPGMKDAVIHALQTSQDTVQCRKASFEL 240
QY 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPVTQAQLCAQVQEDTIKVAV-----DRSCDIGNF 236
DB 241 YGADFVFGVDFOFWLIEINASPTMAPSTAVTARLCAGVQADTLRVVDRRLDRNCDTGAF 300
QY 237 ELLWRQ 242
DB 301 ELIYKQ 306

Search completed: April 4, 2006, 12:58:00
Job time : 34.8606 secs
```

```
RESULT 15
US-10-108-260A-4209
; Sequence 4209, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4209
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-108-260A-4209

Query Match          65.2%; Score 847; DB 4; Length 326;
Best Local Similarity 61.8%; Pred. No. 4.9e-81;
Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1;

QY 1 IDGRLNIWIIPKAAGRGDIVCMRVVEEILEAAADHPLSRDNKVVQKYIETPLLCD 60
DB 61 MEGDRNIWIKPGAKSRGRGIMCWDHLEMLKLVNGNPVVMKDGKVVQKYIERPLIFG 120
QY 61 TKFDIROWFLVTDWNLTIWFYKESYLRFTSTQRFSLKNDLSAHLKNNNAVQKYLKNDVGR 120
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:46:06 ; Search time 9.4461 Seconds
(without alignments)
2118.072 Million cell updates/sec

Title: US-10-635-977-2_COPY_133_374
Perfect score: 1299
Sequence: 1 IDGLRNWIIPAKSRGRD.....IKVAVDRSCDIGNFLLMRQ 242

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/PCITUS_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	271	20.9	439	2	US-10-104-047-3658
2	270	20.8	592	2	US-10-104-047-3371
3	212	16.3	561	2	US-09-270-767-46703
4	177	13.6	404	2	US-10-104-047-3621
5	153.5	11.8	260	2	US-09-270-767-62411
6	153.5	11.8	507	2	US-09-270-767-46787
7	149	11.5	259	2	US-09-248-796A-14603
8	144	11.1	753	2	US-09-538-092-31
9	126	9.7	305	2	US-09-270-767-46577
10	102	7.9	55	2	US-09-270-767-62317
11	91.5	7.0	816	2	US-09-248-796A-20939
12	87.5	6.7	607	2	US-09-489-039A-12140
13	83.5	6.4	268	1	US-07-857-224B-33
14	83	6.4	732	2	US-08-671-757A-7
15	83	6.4	732	2	US-08-671-757A-8
16	83	6.4	732	2	US-09-015-078-7
17	83	6.4	732	2	US-09-015-078-8
18	83	6.4	732	2	US-10-238-977A-7
19	83	6.4	732	2	US-10-238-977A-8
20	80	6.2	243	2	US-09-902-540-15153
21	80	6.2	1577	1	US-08-793-824-2
22	79.5	6.1	433	2	US-09-344-882-26
23	79.5	6.1	433	2	US-10-293-865-26
24	79	6.1	513	2	US-09-351-229-4
25	78.5	6.0	477	2	US-09-075-375F-4
26	78	6.0	256	1	US-08-799-464A-3
27	78	6.0	256	2	US-08-478-316-9

28	78	6.0	256	2	US-08-478-316-14	Sequence 14, Appl
29	78	6.0	256	2	US-09-113-750A-37	Sequence 37, Appl
30	78	6.0	256	2	US-09-019-793A-9	Sequence 9, Appl
31	78	6.0	256	2	US-09-019-793A-14	Sequence 14, Appl
32	78	6.0	256	2	US-09-601-326-9	Sequence 9, Appl
33	78	6.0	256	2	US-09-601-326-14	Sequence 14, Appl
34	78	6.0	256	4	PCT-US95-09927-3	Sequence 3, Appl
35	78	6.0	667	2	US-09-710-279-198	Sequence 198, App
36	77.5	6.0	320	2	US-09-605-703B-808	Sequence 808, App
37	77.5	6.0	320	2	US-09-605-703B-810	Sequence 810, App
38	77.5	6.0	965	2	US-09-252-991A-24038	Sequence 24038, A
39	77	5.9	256	2	US-08-478-316-8	Sequence 8, Appl
40	77	5.9	256	2	US-08-478-316-11	Sequence 11, Appl
41	77	5.9	256	2	US-09-019-793A-8	Sequence 8, Appl
42	77	5.9	256	2	US-09-019-793A-11	Sequence 11, Appl
43	77	5.9	256	2	US-08-301-435-67	Sequence 67, Appl
44	77	5.9	256	2	US-09-601-326-8	Sequence 8, Appl
45	77	5.9	256	2	US-09-601-326-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-10-104-047-3658
; Sequence 3658, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 3658
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3658

Query Match	20.9%	Score 271;	DB 2;	Length 439;
Best Local Similarity	29.8%	Pred. No. 3.8e-22;		
Matches	75;	Conservative 45;	Mismatches 86;	Indels 46;
Gaps	9;			
QY	8	WIIPAKSRGRDIVCMRVEIIELEAAADHPLSRDNK-----WVQKYIETPLLICD	60	
DB	164	WIMKPVARSQKGIFLFRLLKDIVDW-RKDRSSDDQKDDIPVENVAQRYIENPVYIGG	222	
QY	61	TKFDIQWFLVTDNPLTIWFKESYLFSTQFSLDKLSDAIHLCLNNVAKYKLNQVCR	120	
DB	223	RKFDLRVYLVMS-----VFACLLWSGHRR-----QDVHLTNVAVQK-----	260	
QY	121	SPLPAHN-----MWTSTRFOEYL-QROGRGAVGWSVIYPSMKKAIAHAMVAQDHVEPR	174	
DB	261	--TFPDYHPKKGCKWTLQRFQYLAASKHGPEAV--ETLFRDIDNIFVKLSQSVQKVIISD	316	
QY	175	KNSFELYGADPVLGRDFRPWLIEINSSFTMPHTPVTQAQLCAQVOEDTIKVA-----VD	228	
DB	317	KHCFELYGYDILIDQLKPLLEVNASPSLTASSQEDVELKTCLEDTLHVVDMEARLTG	376	
QY	229	RSCDIGNFELMW 240		
DB	377	REKRVGGFDLMW 388		
RESULT 2				
US-10-104-047-3371				
; Sequence 3371, Application US/10104047				
; Patent No. 6943241				
; GENERAL INFORMATION:				

; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46787
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-46787

Query Match 11.8%; Score 153.5; DB 2; Length 507;
Best Local Similarity 35.6%; Pred. No. 1.3e-08;
Matches 37; Conservative 15; Mismatches 39; Indels 13; Gaps 2;

QY 3 GLRNWIIPKAASGRDIVCMRVVEIELEAAADHPLSRDNKWKVQKIETPLLIC--- 59
DB 405 GLDNHWIIPKPNLARGLOTHITDNIKQIVLPAT-----GPKIAQKVIERPVLFSRQE 457

QY 60 ---DTKFDIQWFLVDNPLTIWFKSYLRFSTQRFSLDKLD 100
DB 458 VEGSVKFDIRYVILLKSVKPLKAYTHRKFFLRFANHPFTLDHFD 501

RESULT 7
US-09-248-796A-14603
; Sequence 14603, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14603
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-14603

Query Match 11.5%; Score 149; DB 2; Length 259;
Best Local Similarity 22.9%; Pred. No. 1.6e-08;
Matches 52; Conservative 48; Mismatches 77; Indels 50; Gaps 9;

QY 7 IWIIPKAASGRDI---VCMRVVEIIL-----ELAAADHP-----LSRDNKWV 47
DB 27 LWILKPSMDKGGQIRIFKTLQDQEIFNSFEENDEBEGVDDENGIILSQLRHF 86

QY 48 VQKIETPLLICDTKFDIQWFL---VTDNPLTIWFKSYLRFSTQRFSL-----DKL 99
DB 87 VQYKSNPLLL--SKYDHKKHPLRYYVVCVGDVKVYVKNVLTFLFAGEPYKLPGEDEW 144

QY 100 DSAIHLCCNNAVQKYLKNDVGRSPLLPANHMTSTRFQYLRQGRGAVWGSVIYPSM--- 156
DB 145 SLAGHLNTCLQE-----NEDPLV-----VFFWKLOGLANDKNVIFEQICDI 187

QY 157 -KKAIAHAMKVAQDHVEPRKNSFELYGADFVLGRDPRMLIEINSSP 202
DB 188 TKELFKAAATSDVKNQFQINNAIBIFGVDFLVNSDFSNVLLVNSYP 234

RESULT 8
US-09-538-092-31

; Sequence 31, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratSeqFormatter Version 0.9
; SEQ ID NO 31
; LENGTH: 753
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; NAME/KEY: misc feature
; LOCATION: (0)-(0)
; OTHER INFORMATION: Polypeptide Accession Number YBR094W
US-09-538-092-31

Query Match 11.1%; Score 144; DB 2; Length 753;
Best Local Similarity 21.8%; Pred. No. 2.9e-07;
Matches 57; Conservative 44; Mismatches 105; Indels 56; Gaps 9;

QY 8 WIIPKAASGRDIVCMRVVEI-----LELAADHP----- 39
DB 476 WIVKPSMDKGGQIRVFKTIEDLQAIQFSDDESEAESEGNDDADDVNGEFMDNNKNV 535

QY 40 LSRDNKWKVQKIETPLLIC---DTKFDIQWFLVDNPLTIWFKSYLRFSTQRF--- 94
DB 536 ISQLRHFIQIYLTNPLLASMDNRKKEHRCY--VVCGRGLQVYVDYRMLALFAAKPFVP 593

QY 95 -----SLDKLDSAIHLCCNNAVQKYLKNDVGRSPLLPANHMTSTRFQ--EYLQROGRG 145
DB 594 LDPYAYSVTDLKDLSECHLTNTCLQS-KKKDKDSSVL-----EFDSEIEIENRKS 642

QY 146 AVWGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFELYGADFVLGRDPRMLIEINSSPTMH 205
DB 643 NI-KEQIHSITNDVFLAAVNVNRLNFQPLNAPETYGVDVFLDSDNYEVKLEINAFDPFK 701

QY 206 PSTPTVTAQLCAQVQEDTIKAV 227
DB 702 QTGKDLKKNLIDELFDDTVKYCV 723

RESULT 9
US-09-270-767-46577
; Sequence 46577, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46577
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-46577

Query Match 9.7%; Score 126; DB 2; Length 305;
Best Local Similarity 23.6%; Pred. No. 8.6e-06;
Matches 42; Conservative 36; Mismatches 62; Indels 38; Gaps 8;

QY 48 VQKIETPLLICDTKFDIQWFLVDNPLTIWFKSYL-RFSTQRFSLDKLSAIHL 106

Db 1 VQEFVORPYLVDGKHFIDIGYVYVITSNPLRYIYTGDLVLFYCPVKYHPFDAEN----- 55

Qy 107 NNAVOKYKNDVGRSPLLA-----HNW-----TSTRFOYLQROGR--GAVWGSVI 152

Db 56 ---VDKIIVGD---DYLTWEVPSLRKYNFGSGMRTVFAYVDDQKDPQAIWPQVE 108

Qy 153 Y-----PSMKKAIAHAMVAQDHPVPRKNSFELYGADFLVGRDFFRPLWLEINSSPTM 204

Db 109 HIVRTTIAAKEDIVNLSRYTH-----NFFDLMRFDLFIDEDLKVFLMEANMNSNL 161

RESULT 10

US-09-270-767-62317

; Sequence 62317, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 62317

; LENGTH: 55

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

US-09-270-767-62317

Query Match 7.9%; Score 102; DB 2; Length 55;

Best Local Similarity 43.8%; Pred. No. 0.00038;

Matches 21; Conservative 10; Mismatches 17; Indels 0; Gaps 0;

Qy 47 VVQKIETPLLCITKFDIRQWFLVTDWNLTIWFKESYLRFSTQRF 94

Db 1 VWSKIYVDPCLDGHKCDLRVYLVTSFDPILIYEEGIVRLATVKY 48

RESULT 11

US-09-248-796A-20939

; Sequence 20939, Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 20939

; LENGTH: 816

; TYPE: PRT

; ORGANISM: Candida albicans

US-09-248-796A-20939

Query Match 7.0%; Score 91.5; DB 2; Length 816;

Best Local Similarity 23.7%; Pred. No. 0.33;

Matches 47; Conservative 30; Mismatches 88; Indels 33; Gaps 9;

Qy 20 DIVCMRVEEILEAAADHPLSRDNKVVQKIETPLLCITKFDIRQWFLVTDW---NP 76

Db 454 DELLFLQKSVLQLAICD-MISLDNLV-----LLVHGFTLNRQLFLNCCWYIFNN 504

Qy 77 LTIWFKESYLRFSTQRFSL-----DKLDSAIHLCNNAVOKYKNDVGRSPLPAHNWNT 131

Db 505 LEVLMFAPGFKIEIPNETHQVLEVIDK-----LCN--VQNGTKWIDQNPHLEHLEN 556

Qy 132 STRFOYLQROGRGAVWGSVIYPSMKKAIAHAMVAQ--DHVEPRKNSFELYCADFVLGR 189

Db 557 IEENEYFMDSRRGF---SSFPLDNGYIDAIKTKETVKKRSKRSSTSAFNKDI---S 610

Qy 190 DFRPWLIEINSSPTMHPS 207

Db 611 DFRQGLRRENNGESFPPT 628

RESULT 12

US-09-489-039A-12140

; Sequence 12140, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 12140

; LENGTH: 607

; TYPE: PRT

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-12140

Query Match 6.7%; Score 87.5; DB 2; Length 607;

Best Local Similarity 25.2%; Pred. No. 0.61;

Matches 52; Conservative 23; Mismatches 98; Indels 33; Gaps 11;

Qy 8 WI---IKPAKSGRDIIVCMRVEEILEAAADHPLSRDNKVVQKIETPLLCITKFD 64

Db 175 WWSLLTDPVSGMARDVLSDDAMGYLHFTSGIP--TQGTWML---YSSTFYKMATPPLS 229

Qy 65 -IRQWFLVTDWNL--PLTIWFKESYLRFSTQRFSLDKLDSAIHLCNNAVOKYKNDVGRS 121

Db 230 VINQWQLALDNGSLPAFIAGLAPRHPQYETMHOSL-----LALVADSRPWPQMTGSG 281

Qy 122 PLLPAHNMTST--RFQEVLRQO-----RGAVWGSVIYPSMKKAIAHAMVAQD--V 171

Db 282 SLRPGE--WSNDIGALREILQRTGMLENSANIVLPDGVVSPSAKKSKPAARGVYDRQLV 339

Qy 172 E--PRKNSFELYGADFLVGRDFFRPL 195

Db 340 EGVKRFQAMQGLGADGVIGQSTRDWL 365

RESULT 13

US-07-857-224B-33

; Sequence 33, Application US/07857224B

; Patent No. 5958784

; GENERAL INFORMATION:

; APPLICANT: Benner, Steven A.

; TITLE OF INVENTION: Predicting Folded Structures of Proteins

; NUMBER OF SEQUENCES: 114

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Steven A. Benner

; STREET: Hadlaubstrasse 151

; CITY: Zurich

; STATE: none

; COUNTRY: Switzerland

; ZIP: (note: this is an international post code) CH-8092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage

; COMPUTER: Apple Macintosh

; OPERATING SYSTEM: Macintosh 7.0

; SOFTWARE: Microsoft Word

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/857,224B

; FILING DATE: 03/25/92

CLASSIFICATION: 436
PRIOR APPLICATION DATA: none
TELECOMMUNICATION INFORMATION:
TELEPHONE: (International) 41 1 632 2830
TELEFAX: (International) 41 1 262 2437
TELEX: none
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 268
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: protein
ORIGINAL SOURCE:
ORGANISM: Saccharomyces cerevisiae
FEATURE: Protein kinase; Table 8 Column 37
PUBLICATION INFORMATION:
AUTHORS:
AUTHORS: Hanks, S. K.
AUTHORS: Quinn, A. M.
AUTHORS: Hunter, T.
TITLE: The protein kinase family
JOURNAL: Science
VOLUME: 241
PAGES: 42-52
DATE: 1988
US-07-857-224B-33

Query Match 6.4%; Score 83.5; DB 1; Length 268;
Best Local Similarity 21.4%; Pred. No. 0.52; Mismatches 100; Indels 65; Gaps 11;
Matches 54; Conservative 33;
QY 10 IKPAKSRGRDVCMDRVEEILELAADHP-----LSRDNKWVQKVIETPL----- 56
DB 31 IKEIKTSEFKGLDMSAIREVKYLOEQHPNVIELIDIFMAYDNLNLVLEFLPTLEVI 90
QY 57 ---LLCDTKFDIRQWFLVT-----DWNPLTWYFKESYLRFSQFSLDK 98
DB 91 KDKSLFTPADIKAWMLMTLRGVYHCHRNFLHRLDKPNNLLFSPDQIKVAD--FGLAR 148
QY 99 LDSAIH--LCNNAVOKYLNKD---VGRSPLLPANHMT--STRFOEYLOR---QGRGAVW 148
DB 149 AIPAPHEILTSNVTRWRAPPELLFGAKHYTSADINWSGVIFAEIMLRIPVLPQNDV- 207
QY 149 GSVIYPSMKKAIAHAMKVAQDHVEPRKNSF---ELYGADFVLGRDPRFPWLIEINSSPTWH 205
DB 208 -----DQMEVTFRALGPTDTRDWDPEVSSFWTASEYALDFMCG-----MLTWN 249
QY 206 PSTPVTALCAQ 217
DB 250 PQRWTAQQCLE 261

RESULT 14
US-08-671-757A-7
Sequence 7, Application US/08671757A
Patent No. 6476213
GENERAL INFORMATION:
APPLICANT: Suerbaum, Sebastian
TITLE OF INVENTION: Cloning and Characterization Production
of Aflagellate Strains
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/671,757A
FILING DATE: 16-Aug-2001
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/671,757A
FILING DATE: 16-Aug-2001
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 02356.0073-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 732 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-08-671-757A-7

Query Match 6.4%; Score 83; DB 2; Length 732;
Best Local Similarity 24.5%; Pred. No. 2.6;
Matches 51; Conservative 31; Mismatches 74; Indels 52; Gaps 12;
QY 15 KSRGRDIVCMRVEILELAADHP-LSRDNKWWVQKVIETPL-LICDTKFDIRQWF--- 69
DB 531 KKYAEDFTTKDEVKSLERLARDYPTIVEESKKIPTGAIRSVLQALLHEKIPKMLTTL 590
QY 70 -LVTDMNPL-----TWFYKE-----SYLFEST--QRFSLDKLSA 102
DB 591 ETITDIALVQNDVNILTEQVRLSRVITNAFSEDGRLKELTSTDEQPLLNKLR-- 648
QY 103 IHLCNNAVOKYLNKDVR-SPLLPAHNMTSTRFOEYLORQGRGAVMSVYIYPSMKKAIA 161
DB 649 ---ENGTSKSLNVLNGELQKLEA---VSEEMKVLQKGIAPVI--LIVEPNLRKALS 698
QY 162 HAMKVAQ-----DHVEPRKNS-PELYG 182
DB 699 NOMEQARIDIVLSHAELDPNSFEALG 726

RESULT 15
US-08-671-757A-8
Sequence 8, Application US/08671757A
Patent No. 6476213
GENERAL INFORMATION:
APPLICANT: Suerbaum, Sebastian
TITLE OF INVENTION: Cloning and Characterization Production
of Aflagellate Strains
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/671,757A
FILING DATE: 16-Aug-2001
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146

```

;
; REFERENCE/DOCKET NUMBER: 02356.0073-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 732 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-08-671-757A-8

Query Match      6.4%; Score 83; DB 2; Length 732;
Best Local Similarity 24.5%; Pred. No. 2.6;
Matches 51; Conservative 31; Mismatches 74; Indels 52; Gaps 12;

QY 15 KSRGRDIVCMRVVEEILELAAADHP-LSRDNKVVQVQYIETPL-LICDTKFDIRQWF--- 69
Db 531 KKYAEDFITKDEVKSLERLAKDYPTIVEESKKIPTGAIRSVLQALLHEKIPKIDMLTIL 590
QY 70 -LVTDNPL-----TIFVKE----SYLRFST--ORFSLDKLDSA 102
Db 591 ETITDIAPLVQNDVNILTEQVRARLSRVITNAFKSEGDGRKLFITFTDSEQFLNKLRL-- 648
QY 103 IHLCNNAVQKYLKNDVGR-SPLLPFAHNMWTTSTFQEYLQQRGAVGWSVIYPSMKKAIA 161
Db 649 -----ENGTSKSLNLVNGELQKLIEA---VSEEMKVLQKGIAPVI--LIVEPNLRKALS 698
QY 162 HAMKVAQ-----DHVEPRKNS-FELYG 182
Db 699 NOMEQARIDIVILSHAELDPNSNFALG 726
```

Search completed: April 4, 2006, 12:46:56
Job time : 9.4461 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:43:06 ; Search time 36.6599 Seconds
(without alignments)
4657.352 Million cell updates/sec

Title: US-10-635-977-2_COPY_133_374
Perfect score: 1299
Sequence: 1 IDGLRNIIWIKPAKSRGRD.....IKVAVDRSCDIGNFELLWRQ 242

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1287	99.1	747	2	Q5JZ84_HUMAN	Q5JZ84 homo sapien
2	1143	88.0	518	2	Q8C0V2_MOUSE	Q8C0V2 mus musculus
3	1143	88.0	781	2	Q8CON7_MOUSE	Q8CON7 mus musculus
4	850	65.4	352	1	TTLL3_HUMAN	Q9Y4R7 homo sapien
5	850	65.4	352	2	Q6AWA3_HUMAN	Q6AWA3 homo sapien
6	850	65.4	352	2	Q4KMS8_HUMAN	Q4KMS8 homo sapien
7	850	65.4	434	2	Q8NDN8_HUMAN	Q8NDN8 homo sapien
8	850	65.4	744	2	Q9H876_HUMAN	Q9H876 homo sapien
9	838	64.5	704	2	Q8BV51_MOUSE	Q8BV51 mus musculus
10	834	64.2	266	2	Q922T0_MOUSE	Q922T0 mus musculus
11	799	61.5	261	2	Q58CT2_BOVIN	Q58CT2 bos taurus
12	786	60.5	534	2	Q4RY08_TENG	Q4RY08 tetraodon n
13	644.5	49.6	572	2	Q7Q156_ANOGA	Q7Q156 anopheles g
14	628.5	48.4	992	2	Q9VM91_DROME	Q9VM91 drosophila
15	521	40.1	501	2	Q7PMD3_ANOGA	Q7PMD3 anopheles g
16	499.5	38.5	719	2	Q9VM92_DROME	Q9VM92 drosophila
17	499.5	38.5	756	2	Q5BHY1_DROME	Q5BHY1 drosophila
18	468	36.0	281	2	Q6ZU95_HUMAN	Q6ZU95 homo sapien
19	423	32.6	331	2	Q5TN29_ANOGA	Q5TN29 anopheles g
20	385	29.6	101	2	Q9GGG8_HUMAN	Q9GGG8 homo sapien
21	342	26.3	461	2	Q641W7_RAT	Q641W7 rattus norv
22	342	26.3	464	2	Q9D570_MOUSE	Q9D570 mus musculus
23	340.5	26.2	523	2	Q54TJ1_DICDI	Q54TJ1 dictyosteli
24	328.5	25.3	794	2	Q6BFH6_PARTE	Q6BFH6 paramescium
25	324	24.9	496	2	Q961I9_DROME	Q961I9 drosophila
26	324	24.9	496	2	Q9VX74_DROME	Q9VX74 drosophila
27	320.5	24.7	1075	2	Q7QT13_GIALA	Q7QT13 giardia lam
28	313	24.1	341	2	Q5VX47_HUMAN	Q5VX47 homo sapien
29	299	23.0	1339	2	Q5F498_CHICK	Q5F498 gallus gall
30	297.5	22.9	827	2	Q9VKL9_DROME	Q9VKL9 drosophila
31	297.5	22.9	828	2	Q8IGW4_DROME	Q8IGW4 drosophila

RESULT 1

Q5JZ84_HUMAN
AC Q5JZ84_HUMAN PRELIMINARY; PRT; 747 AA.
DT 10-MAY-2005 (Tremblrel. 30, Created)
DT 10-MAY-2005 (Tremblrel. 30, Last sequence update)
DE OTTHUMP0000028514 (Fragment).
GN ORFNames=RP3-355C18.2-002;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]_TTL; 1.
RP NUCLEOTIDE SEQUENCE.
RA Cobley V.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022327; CAI42686.1; -; Genomic DNA.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_lygase.
DR Pfam; PF03133; TTL; 1.
FT NON_TER 1
SQ SEQUENCE 747 AA; 83854 MW; F90948E159BBE589 CRC64;

Query Match

Best Local Similarity 99.1%; Score 1287; DB 2; Length 747;
Matches 242; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 IDGLRNIIWIKPAKSRGR-----DIVCNDRVVEEILEAAADHPLSRDNKVVQKYIETPL 56
DB 335 IDGLRNIIWIKPAKSRGRGSDIVCNDRVVEEILEAAADHPLSRDNKVVQKYIETPL 394
QY 57 LICDTKFDIRQWFLVTDNPLTIWPKSYLRFSTQSFSLDKLSAIHLCNNAVQKYLKN 116
DB 395 LICDTKFDIRQWFLVTDNPLTIWPKSYLRFSTQSFSLDKLSAIHLCNNAVQKYLKN 454
QY 117 DVGSRPLPAHNMWTSFQVYLOQRGAVGWSVIYPSMKKAI AHAMKVAODHVEPRKN 176
DB 455 DVGSRPLPAHNMWTSFQVYLOQRGAVGWSVIYPSMKKAI AHAMKVAODHVEPRKN 514
QY 177 SFELYGADFVLGRDPRPMLIENSPTMHPSTPTVTAQLCAOVQEDTIKAVDRSCDIGNF 236
DB 515 SFELYGADFVLGRDPRPMLIENSPTMHPSTPTVTAQLCAOVQEDTIKAVDRSCDIGNF 574
QY 237 ELLWRQ 242
DB 575 ELLWRQ 580

RESULT 2

Q8C0V2_MOUSE
ID Q8C0V2_MOUSE PRELIMINARY; PRT; 518 AA.

AC Q8C0V2; NUCLEOTIDE SEQUENCE.
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Mus musculus adult male testis cDNA, RIKEN full-length enriched
 DE library, clone:4930524K07 product:hypothetical Tubulin-tyrosine ligase
 DE containing protein, full insert sequence.
 DE Name:1700019P01Rik;
 GN Mus musculus (Mouse).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridea; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Testis;
 RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RX Carninci P., Hayashizaki Y.;
 RA "High-efficiency full-length cDNA cloning";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Testis;
 RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RX Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Scahill F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RL "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690(2001).
 RN [3]
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Testis;
 RC The FANTOM Consortium;
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573(2002).
 RN [4]
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Testis;
 RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RX Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Testis;
 RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RX Shibata K., Itoh M., Aizawa K., Nagasaki S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa H., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RL "RIKEN integrated sequence analysis (RISA) system-384-format

RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Testis;
 RC Adachi J., Aizawa K., Akimura T., Hara A., Hashizume W.,
 RA Fukuda S., Furuno M., Hanagaki T., Hiraoka T., Hirozane T.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Kagawa I., Kasukawa T.,
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa T., Koyama T.,
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK029745; BC26595.1; -; mRNA.
 DR Ensembl; ENSMUSG00000022388; Mus musculus.
 DR MGI; MGI:1922902; 1700019P01Rik;
 DR GO; GO:0016874; F.ligase activity; IEA.
 DR GO; GO:0004835; F.tubulin-tyrosine ligase activity; IEA.
 DR GO; GO:0006464; P.protein modification; IEA.
 DR InterPro; IPR004344; Tub_tyr_ligase.
 DR Pfam; PF03133; TTL; 1.
 KW Hypothetical protein; Ligase.
 SQ SEQUENCE 518 AA; 59319 MW; 540C045659FAC0D6 CRC64;
 Query Match 88.0%; Score 1143; DB 2; Length 518;
 Best Local Similarity 85.0%; Pred. No. 5.4e-94;
 Matches 209; Conservative 17; Mismatches 16; Indels 4; Gaps 1;
 QY 1 IDGRNIWIIPKPAKGRDIVCMWDVEETLELAADHPLSRDNKVVQVQKIETPLICD 60
 DB 189 IDGRNIWIIPKPAKGRDIVCMWDVEETLELAADHPLSRDNKVVQVQKIETPLIYD 248
 QY 61 TKFDIQWFLVTDNPLTIWFKYESYRSTQFSLDKLDSAIHLNNAVQKYLKNDVGR 120
 DB 249 TKFDIQWFLVTDNPLTIWFKYESYRSTQFSLDKLDSAIHLNNSIQRLKNDKER 308
 QY 121 SPLLPAAHNMWTRFQYELQQRGVGWSVYPSMKKATAHAMKVAQDHVEPRKNSFEL 180
 DB 309 SPLLPCHNMWTRFQYELQQRGVGWSVYPSMKKRAVTNMRVAQDHVEARKNSFEL 368
 QY 181 YGADFVLGRDPRFWLIEINSSPTMHSTPTVTAQLCAQVQEDTIKVAV----DRSCDIGNF 236
 DB 369 YGADFVLGRDPRFWLIEINSSPTMHSTPTVTAQLCAQVQEDTIKVAVVDRKLDRCNDIGNF 428
 QY 237 ELLWRQ 242
 DB 429 ELLWRQ 434
 RESULT 3
 Q8CON7_MOUSE PRELIMINARY; PRT; 781 AA.
 ID Q8CON7_MOUSE PRELIMINARY; PRT; 781 AA.
 AC Q8CON7;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Mus musculus adult male testis cDNA, RIKEN full-length enriched
 DE library, clone:4933401B17 product:hypothetical Tubulin-tyrosine ligase
 DE containing protein, full insert sequence.
 DE Name:1700019P01Rik;
 GN Mus musculus (Mouse).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridea; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Testis;
 RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RX Carninci P., Hayashizaki Y.

RA Carninci P., Hayaishizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuoka H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Giasi C., King B., Koichiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.P.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayaishizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX The FANTOM Consortium,
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayaishizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayaishizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanaoka T., Hara A., Hashizume W.,
 RA Hayaishida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nihi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akaira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayaishizaki Y.;
 RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AK030151; BAC26811.1; -, mRNA.
 DR Ensembl; ENSMUSG0000022388; Mus musculus.
 DR MGI; MGI:1922902; 170019P01Rik.
 DR GO; GO:0016874; P:ligase activity; IEA.
 DR GO; GO:0004835; P:tubulin-tyrosine ligase activity; IEA.
 DR GO; GO:0006464; P:protein modification; IEA.
 DR InterPro; IPR000276; GPCR_Rhodopsin.
 DR Pfam; PF03133; TTL; 1.
 DR PROSITE; PS00237; G_PROTEIN_RECEPTOR_1; UNKNOWN_1.
 KW Hypothetical protein; Ligase.
 SQ SEQUENCE 781 AA; 89398 MW; 9E98793C3351C3DE CRC64;
 Query Match 88.0%; Score 1143; DB 2; Length 781;
 Best Local Similarity 85.0%; Pred. No. 8.8e-94;
 Matches 209; Conservative 17; Mismatches 16; Indels 4; Gaps 1;
 QY 1 IDGLNINIIIPAAKSRGRDIVCMRDVEILELAADHPDLSRDNKKVQKVIETPLLCD 60
 DB 336 IDGIENIWIIPAAKSRGRDIVCMRDVENILSLVAADSQTDXDNKVVQKVIETPLIYD 395
 QY 61 TKFDIROWFLVTDNPLTIWFYKESYLSFSTORFSLDKLDSAIHLCLNNVAVOKYLVNDVGR 120
 DB 396 TKFDIROWFLVTDNPLTIWFYKESYLSFSTORFSLDKLDSAIHLCLNNVAVOKYLVNDVGR 455
 QY 121 SPLLPANHWITSTRFEYLQRCGAVMGSVIYPSMKKAIAMKVAODHVEPRKNSFEL 180
 DB 456 SPLLPANHWITSTRFEYLQRCGAVMGSVIYPSMKKAIAMKVAODHVEPRKNSFEL 515
 QY 181 YGADPVLGRDPRPMLIENSSPTMHPSTPTVAQLCAQVOEDTIKAVV-----DRSCDIGNF 236
 DB 516 YGADPVLGRDPRPMLIENSSPTMHPSTPTVAQLCAQVOEDTIKAVV-----DRSCDIGNF 575
 QY 237 ELLWRQ 242
 DB 576 ELLWRQ 581

RESULT 4
 ID TTL3 HUMAN
 TT TTL3 HUMAN STANDARD; PRT; 352 AA.
 AC Q94R7; Q9U199;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)
 DE Tubulin tyrosine ligase-like protein 3 (HOTTTL).
 GN Name=TTL3; ORFNames=PRO0207;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RC TISSUE=Testis;
 RG The German cDNA consortium;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 61-352.
 RC TISSUE=Fetal liver;
 RA Yu Y., Zhang C., Luo L., Ouyang S., Zhang S., Li W., Wu J., Zhou S.,
 RA Liu M., He F.;
 RT "Functional prediction of the coding sequences of 50 new genes deduced
 by analysis of cDNA clones from human fetal liver.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Contains 1 TTL domain.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use as long as its content is in no way modified and this statement is not
 removed.

DR	EMBL; AL833939;	CAD38794.1;	-;	mRNA.
DR	EMBL; BC098298;	AAH98298.1;	-;	mRNA.
DR	GO; GO:0004835;	F:tubulin-tyrosine ligase activity;	IEA.	
DR	GO; GO:0006464;	P:protein modification;	IEA.	
DR	InterPro; IPR004344;	Tub_tyr_ligase.		
DR	Pfam; PF03133;	TTL; 1.		
DR	KW	Hypothetical protein.		
SQ	SEQUENCE	434 AA;	94933 MW;	957956CA08651CAl CRC64;

Query Match	65.4%;	Score 850;	DB 2;	Length 434;
Best Local Similarity	61.8%;	Pred. No. 9.7e-68;		
Matches 152;	Conservative 41;	Mismatches 49;	Indels 4;	Gaps 1;

Oy	1	IDGRLNTWIILKPAAKSRGRDVCMDRVVEETLELAADPHLSRDNKNVVKYIETPLLICD	60
Dd	61	MEGRDNIWIKPKAGSKRGIMCDMLHEMLKLGVNGPVVKGKGVVVQKIETRPLLIIFG	120
Oy	61	TKFDIRQWLFTDWNPTITFIYKESYLRFSTQFSLDKLSDAIHLCKNNAVQKYLNKDVGVR	120
Dd	121	TKFDLRQWLFTDWNPLTWVFYRDSYIRFSTQPFSLKNDLNSVHLCKNNSIQKHLENSCHR	180
Oy	121	SPLLPAHNMTWTSTRFOEYLQRQGRGAVGWGSVIYPSSKKATAAHAMKVQAODHVPRKNSFEL	180
Dd	181	HPLLPPDNWSSQRFQAHLOEQMGNAPNAWSIIIVPGMKDAVIHALQTSQDTVQCRCASFEL	240
Oy	181	YGADFVLGRDFRWLLEINSSPWHPTSTPYTAOLCAQVQEDTTIKVAV---	236
Dd	241	YGADFVGEDFPQLWELEINASPMAWSTATVARTLCAGVQADTURVDRLDRNCDTGAF	300
Oy	237	ELLWRQ 242	
Dd	301	ELIVKQ 306	

RESULT 8			
Q9RH876 HUMAN			
ID	Q9RH876 HUMAN PRELIMINARY;	PRT;	744 AA.
AC	Q9RH876;		
CD	01-MAR-2001 (TrEMBLrel. 16, Created)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)		
DE	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Hypothetical protein FLJ13898.		
OS	Homo sapiens (human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;		
OC	Homo.		
OX	NCBI_TaxID=9606;		
[1]	NUCLEOTIDE SEQUENCE.		
RN	TISSUE=Thyroid gland;		
RC	PubMed=14702039; DOI=10.1038/ngl1285;		
RA	Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R., Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H., Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S., Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahori K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ihida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J., Kanehori K., Takahashi-Fuji A., Hara H., Tanase T.-O., Nomura Y., Togaya S., Komai F., Hara R., Takeuchi K., Arima M., Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S., Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S., Moriwa S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O., Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu Y., Wakebe H., Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B., Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y., Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T., Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,		

RA	Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,	
RA	Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,	
RA	Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,	
RA	Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,	
RA	Okumura K., Nagase T., Nomura N., Kikuchi H., Maeuho Y., Yamashita R.,	
RA	Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;	
RT	"Complete sequencing and characterization of 21,243 full-length human	
RT	cDNAs."	
RL	Nat. Genet. 36:40-45(2004).	
RL	EMBL; AK023960; BAB14741.1; -; mRNA.	
DR	Ensembl; ENSG00000156983; Homo sapiens.	
DR	GO; GO:0016874; F:ligase activity; IEA.	
DR	GO; GO:0004935; F:tubulin-tyrosine ligase activity; IEA.	
DR	GO; GO:0006464; P:protein modification; IEA.	
DR	InterPro; IPR004344; Tub_tyr_lygase.	
DR	Pfam; PF03133; TTL; 1.	
KW	Ligase.	
FT	NON TER	744 744
SQ	SEQUENCE	744 AA; 84683 MW; DF661753E4AFF0DF CRC64;
	Query Match	65.4%; Score 850; DB 2; Length 744;
	Best Local Similarity	61.8%; Pred. No. 1.8e-67;
	Matches 152; Conservative	41; Mismatches 49; Indels 4; Gaps 1;
Qy	1	IDGLRNTWIIIPKAASGRDIVCMRDVVEILELAADHPISRDNNKVVQVQKIETPLLICD 60
Db	273	MEGRDNIWIKPGAKSRGRGIMCMDHLEMLKLVGNPNVVMKDGKVVQVQKIETPLLI 332
Qy	61	TKPDIROWFLVTDNPLTIWFYKESYLRFSTQRESLKDLSAHLCNNAVQYKLVNDVGR 120
Db	333	TKFDLRQWFLVTDNPLTVWFYRDSYTRFSTQPFSLKVLNDSVHLCNNSIQKHLENSCHR 392
Qy	121	SPLLPAHNMWTSRTFOEYLQSGRGAVNGSVIYPSMKKAIHAHMKVADQHVPRKNSPEL 180
Db	393	HLPLPPDNWSSQRFQAHLEQMGAPNAWSTIIIVFGMKDAVIAHQTSDTVOCKKASEL 452
Qy	181	YGADFVLGRDRFPWLIEINSPTMHPSTPTVAQLCAQVOEDTIKVAV---DRSCDIGNF 236
Db	453	YGADFVFGEDFPWLIEINASPTMAPSTAVTARLCAGVQADTLRVVIDRMLDRNCDTGAF 512
Qy	237	ELLWRQ 242
Db	513	ELIYKQ 518
RESULT 9		
Q8BV51_MOUSE		
ID	Q8BV51_MOUSE PRELIMINARY;	PRT; 704 AA.
AC	Q8BV51;	
DT	01-MAR-2003 (TrEMBLrel. 23, Created)	
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)	
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE	Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched	
DE	library, clone:A630053H17 product:HOTTI, PROTEIN homolog.	
GN	Name=4833441J24Rik;	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;	
OC	Muroidea; Muridae; Murinae; Mus.	
OX	NCBI_TaxID=10090;	
ON	[1]	
RP	NUCLEOTIDE SEQUENCE.	
RC	STRAIN=C57BL/6J; TISSUE=Thymus;	
RX	MEDLINE=92279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;	
RA	Carninci P., Hayashizaki Y.;	
RA	"High-efficiency full-length cDNA cloning.";	
RL	Meth. Enzymol. 303:19-44(1999).	
RL	[2]	
RP	NUCLEOTIDE SEQUENCE.	
RC	STRAIN=C57BL/6J; TISSUE=Thymus;	
RX	MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;	
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,	
RA	Kawakawa T., Hara A.,	

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saïto T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fletschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Bono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustingich S., Hall D., Hofmann C., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momboerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weiz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohteuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RC THE FANTOM Consortium,
RC the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RT Nature 420:563-573(2002).
RL [4]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RC Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RT Genome Res. 10:1617-1630(2000).
RL [5]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RC MEDLINE=20350913; PubMed=11076863; DOI=10.1101/gr.152600;
RC Shibata K., Itoh M., Aizawa K., Nagakura S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama N., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RL [6]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RC Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Tasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akaira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK080321; BAC37878.1; -; mRNA.
RL Ensembl; ENSMUSG0000030276; Mus musculus.
RL MGI; MGI:2141418; 4833441J24Rik.
DR GO; GO:0004835; F-ubiquitin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P-protein modification; IEA.
DR Interpro; IPR004344; Tub_tyrligase.
DR Pfam; PF03133; TUB1; 1.
SQ SEQUENCE 703 AA; 79080 MW; 3FAD899C1DB5CF7D CRC64;

Query Match 64.5%; Score 838; DB 2; Length 704;
Best Local Similarity 61.0%; Pred. No. 2.1e-66;
Matches 150; Conservative 43; Mismatches 49; Indels 4; Gaps 1;

QY 1 IDGLRNIIWIKPAKSRGRDIVCMRDVBEILEAAADHPLSRDNKVVQKVIETPLLICD 60
DB 243 MEGDRNIWIVPGAKSRGRGIMCMNRLDMLKLVDCNPNMLKDGKWIQVQKVIETPLIFG 302

QY 61 TKFDTRQWFLVTDWNPFLTTFWFKESYLRFPSTQRFSLDKLDSAIHLNNAVQKYLKNDVGR 120
DB 303 TKFDLRQWFLVTDWNPFLTTFWFKESYLRFPSTQRFSLDKLDSVHLNNSIQRLHLEASCHR 362

QY 121 SPLLPAHNMWTSRFOEYLQRGGRGAVGWSVIVPSMKKAIAHAMKVAODHVPKRNKSPFL 180
DB 363 HPMLPPDNWSSQRFQAHLEQVDAPKAWSSVIVPGKAAVIAHALQTSQDNVQCRKASFEL 422

QY 181 YGADFVFLGRDFPWLIEINSSPTMHPSTVTAQLCAQVQEDTIKAVAV-----DRSCDIGNF 236
DB 423 YGADFVFGEDFQWLIENASPTMAPSTAVTARLCAGVQADTLRVVDRDRSDCTGAP 482

QY 237 ELLWRQ 242
DB 483 ELIYKQ 488

RESULT 10
Q922T0 MOUSE
ID Q922T0_MOUSE PRELIMINARY; PRT; 266 AA.
AC Q922T0;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE 4833441J24Rik;
GN Name=4833441J24Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RX MEDLINE=23388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Trichman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinchwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RG NIH MGC Project;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006830; AA06830.1; -; mRNA.
DR Ensembl; ENSMUSG00000030276; Mus musculus.
DR MGI; MGI:2141418; 4833441J24Rik.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.

DR InterPro; IPR004344; Tub_tyr_ligase.
DR Pfam; PF03133; TTL; 1.
SQ SEQUENCE 266 AA; 30506 MW; F8B8F852FA8B8E98 CRC64;

Query Match 64.2%; Score 834; DB 2; Length 266;
Best Local Similarity 60.6%; Pred. No. 1.5e-66;
Matches 149; Conservative 44; Mismatches 49; Indels 4; Gaps 1;

QY 1 IDGLRNIIWIKPAKSRGRDIVCMRDVBEILEAAADHPLSRDNKVVQKVIETPLLICD 60
DB 1 MEGDQNIWIVPGAKSRGRGIMCMNRLDMLKLVDCNPNMLKDGKWIQVQKVIETPLIFG 60

QY 61 TKFDTRQWFLVTDWNPFLTTFWFKESYLRFPSTQRFSLDKLDSAIHLNNAVQKYLKNDVGR 120
DB 61 TKFDLRQWFLVTDWNPFLTTFWFKESYLRFPSTQRFSLDKLDSVHLNNSIQRLHLEASCHR 120

QY 121 SPLLPAHNMWTSRFOEYLQRGGRGAVGWSVIVPSMKKAIAHAMKVAODHVPKRNKSPFL 180
DB 121 HPMLPPDNWSSQRFQAHLEQVDAPKAWSSVIVPGKAAVIAHALQTSQDNVQCRKASFEL 180

QY 181 YGADFVFLGRDFPWLIEINSSPTMHPSTVTAQLCAQVQEDTIKAVAV-----DRSCDIGNF 236
DB 181 YGADFVFGEDFQWLIENASPTMAPSTAVTARLCAGVQADTLRVVDRDRSDCTGAP 240

QY 237 ELLWRQ 242
DB 241 ELIYKQ 246

RESULT 11
Q98CT2 BOVIN
ID Q98CT2_BOVIN PRELIMINARY; PRT; 261 AA.
AC Q98CT2;
DT 10-MAY-2005 (TREMBlrel. 30, Created)
DT 10-MAY-2005 (TREMBlrel. 30, Last sequence update)
DT 10-MAY-2005 (TREMBlrel. 30, Last annotation update)
DE Tubulin tyrosine ligase-like family, member 3.
GN Name=TTLL3;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC Pecora; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Pooled;
RX MEDLINE=21180013; PubMed=11282978; DOI=10.1101/gr.170101;
RA Smith T.P.L., Grosse W.M., Freking B.A., Roberts A.J., Stone R.T.,
RA Casas E., Wray J.E., White J., Cho J., Fahrenkrug S.C., Bennett G.L.,
RA Heaton M.P., Laegreid W.W., Rohrer G.A., Chitko-McKown C.G.,
RA Petrea G., Holt I., Karamycheva S., Liang F., Quackenbush J.,
RA Keele J.W.;
RT "Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle."
RL Genome Res. 11:626-630 (2001).
[2]
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Pooled;
RA Hathay G.P., Sonstegard T.S., Van Tassell C.P., Clawson M.L.,
RA Heaton M.P., Keele J.W., Snelling W.M., Weidmann R.T., Smith T.P.L.;
RT "Sequencing and analysis of Bos taurus full-length insert cDNA
clones."
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; BT021865; AAX46712.1; -; mRNA.
KW Ligase.
SQ SEQUENCE 261 AA; 29911 MW; 5F069784CA162017 CRC64;

Query Match 61.5%; Score 799; DB 2; Length 261;
Best Local Similarity 61.6%; Pred. No. 2e-63;
Matches 143; Conservative 37; Mismatches 52; Indels 0; Gaps 0;

QY 1 IDGLRNIIWIKPAKSRGRDIVCMRDVBEILEAAADHPLSRDNKVVQKVIETPLLICD 60
DB 1 MEGDQNIWIVPGAKSRGRGIMCMNRLDMLKLVDCNPNMLKDGKWIQVQKVIETPLIFG 60

```
Db 3 MEGDRNIWIVKPGAKSRGRCIMCWDHLEBMLKLVDNCNPMWKGKWWVHNYIERPLLIIG 62
Qy 61 TKFDIRQWFLVTDWNPPLTTFWFKESYLRSTQRFSLDKLDSAIHLCCNAVOKYKLVNDVGR 120
Db 63 TKFDLRQWFLVTDWNPPLTTFWFKESYLRSTQRFSLDKLDSAIHLCCNAVOKYKLVNDVGR 122
Qy 121 SPLLPAHNNWTSRFOEYLQROGRGAVGWSVLYPSMKKAIAHAMKVAQDHVPRKNSFEL 180
Db 123 HPLLPPDNWSSQKFOAHQETGAPNAWSTVIVPGMKAAVIFALQTSQDTVQCCKASFEL 182
Qy 181 YGADFVLGRDPRFWLLEINSSPTMHPSTPTVTAOLCAQVQEDTIKVAVDKSCD 232
Db 183 YGADFVFGEDFQWLEINASSPTMASTAVTARLCAGVQADTLRVVIDWRLD 234

RESULT 12
Q4RY08_TETNG PRELIMINARY; PRT; 534 AA.
AC Q4RY08_TETNG PRELIMINARY; PRT; 534 AA.
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Chromosome 11 SCAFI4979, whole genome shotgun sequence.
GN ORFNames=GSTENG0027209001;
OS Tetraodon nigroviridis (green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cottarel L., Poulain J., De Berardinis V.,
RA Craud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Queirer F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RL the early vertebrate proto-karyotype.";
RN Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Genoscope; Whitehead Institute Centre for Genome Research;
RA Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAE01014979; CAG06724.1; -; Genomic DNA.
DR EMBL; 534 AA; 60130 MW; 62B5B5B696B62B CRC64;
SQ SEQUENCE 534 AA; 60130 MW; 62B5B5B696B62B CRC64;

Query Match 60.5%; Score 786; DB 2; Length 534;
Best Local Similarity 59.2%; Pred. No. 7e-62;
Matches 145; Conservative 36; Mismatches 60; Indels 4; Gaps 1;

Qy 2 DGLNIIWIKPAKSRGRDIVCMRDVBEILELAADHPLSRDNKVVQKYIETPLLCDT 61
Db 262 DGLNIIWIKPAKSRGRDIVCMRDVBEILELAADHPLSRDNKVVQKYIETPLLCDT 321
Qy 62 KFDTRQWFLVTDWNPPLTTFWFKESYLRSTQRFSLDKLDSAIHLCCNAVOKYKLVNDVGRS 121
Db 322 KFDTRQWFLVTDWNPPLTTFWFKESYLRSTQRFSLDKLDSAIHLCCNAVOKYKLVNDVGRS 381
Qy 122 PLLPAHNNWTSRFOEYLQROGRGAVGWSVLYPSMKKAIAHAMKVAQDHVPRKNSFELY 181
Db 382 PAVPGDNWSSQKFOAHQETGAPNAWSTVIVPGMKAAVIFALQTSQDTVQCCKASFELY 441
```

```
Qy 182 GADFVLGRDPRFWLLEINSSPTMHPSTPTVTAOLCAQVQEDTIKVAVDKSCD----IGNFE 237
Db 442 GADFVLGRDPRFWLLEINSSPTMHPSTPTVTAOLCAQVQEDTIKVAVDKSCD----IGNFE 501
Qy 238 LLWRQ 242
Db 502 LLWRQ 506

RESULT 13
Q7Q156 ANOQA PRELIMINARY; PRT; 572 AA.
AC Q7Q156;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE ENSANGP0000022337 (Fragment).
GN ORFNames=ENSANG0000019848;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
OC Anophelinae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RT "Anopheles gambiae re-annotation.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008980; EAA13905.2; -; Genomic DNA.
DR GO; GO:0004835; Fubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; Protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_lygase.
DR Pfam; PF03113; TTL; 1.
FT NON_TER 1 572
FT NON_TER 572 572
SQ SEQUENCE 572 AA; 67109 MW; 8938D9EDD5935071 CRC64;

Query Match 49.6%; Score 644.5; DB 2; Length 572;
Best Local Similarity 50.4%; Pred. No. 4e-49;
Matches 124; Conservative 45; Mismatches 70; Indels 7; Gaps 3;

Qy 1 IDGLRNIWIKPAKSRGRDIVCMRDVBEILELAADHPLSRDNKVVQKYIETPLLCD 60
Db 305 LDGLRNIWIKPAKSRGRDIVCMRDVBEILELAADHPLSRDNKVVQKYIETPLLCD 362
Qy 61 TKFDIRQWFLVTDWNPPLTTFWFKESYLRSTQRFSLDKLDSAIHLCCNAVOKYKLVNDVGR 120
Db 363 TKFDIRQWFLVTDWNPPLTTFWFKESYLRSTQRFSLDKLDSAIHLCCNAVOKYKLVNDVGR 421
Qy 121 SPLLPAHNNWTSRFOEYLQROGRGAVGWSVLYPSMKKAIAHAMKVAQDHVPRKNSFEL 180
Db 422 DERLPENWDCHTFQAYLRQIDKYENWSEIRIYVPGMKAIGSILLACQDNDRPNTFEL 481
Qy 181 YGADFVLGRDPRFWLLEINSSPTMHPSTPTVTAOLCAQVQEDTIKVAVDKSCD----IGNF 236
Db 482 YGADFVITTFWFLLEINSSPTLAFSTSVTARLCPCQVEDTIRVIDRRTDSNAPTGSF 541
Qy 237 LLWRQ 242
Db 542 ELIYKQ 547

RESULT 14
Q9VM91_DROME
```


ID Q9VM91_DROME PRELIMINARY; PRT; 992 AA.
 AC Q9VM91;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CG11323-PA.
 GN Name=CG11323; ORFNames=CG11323;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1] _
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Ballieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burcis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwac C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko F., Lei Y., Levitsky A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphay L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle B.J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195 (2000).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
 RA Pacle B.J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*
 RT melanogaster euchromatic genome sequence.";
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirskas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.;
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
 RT a genomics perspective."

RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20 (2002).
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).
 RN [5]
 RP NUCLEOTIDE SEQUENCE.
 RG Berkeley Drosophila Genome Project;
 RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
 RA Hoskins R., Stapleton M., Pacle B.J., Park S., Svirskas R., Smith E.,
 RA Yu C., Rubin G.;
 RT "Drosophila melanogaster release 4 sequence.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP NUCLEOTIDE SEQUENCE.
 RG FlyBase;
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003614; AAF52432.1; -; Genomic DNA.
 DR Ensembl; CG11323; Drosophila melanogaster.
 DR FlyBase; FBgn0031854; CG11323.
 DR GO; GO:0004835; F: tubulin-tyrosine ligase activity; IEA.
 DR GO; GO:0006464; P: protein modification; IEA.
 DR InterPro; IPR004344; Tub_tyr_lygase.
 DR Pfam; PF03133; TTL; 1.
 SQ SEQUENCE 992 AA; 112255 MW; 0FC34674C15BDAAD CRC64;
 Query Match 48.4%; Score 628.5; DB 2; Length 992;
 Best Local Similarity 48.4%; Pred. No. 2.1e-47;
 Matches 119; Conservative 50; Mismatches 70; Indels 7; Gaps 3;
 QY 1 IDGNINWIKPAKSRGRDVCMDRVEELLELAADHPLSDKNVVKYIETPLICD 60
 DB 411 LDGQNMWIVKPAKRCGRGIILMDNKKI--LGVVNLISIAKSRVYVQKIERPLILFQ 468
 QY 61 TKFDIOWFLVTDNPLTIWFYKESYLFSTQFSLDKLDSAIHLCCNNAVKYLKNDVGR 120
 DB 469 TKFDIOWFLITNTQPLVVMYRESYLFSSQESYLSNHHSHVLTNTVAIQKYNG-KR 527
 QY 121 SPLPAHNMWTSRFOEYLRQGRGAVMGSVIYPSMKKIAAHAMKVAQDHVEPRKNSPEL 180
 DB 528 DKRLPSENMDVCSFYALRQIGKYNMWERIFPGMKRAIVGCMLASQENMDRRPNTPEL 587
 QY 181 YGADVFLGRDPRPWLIEINSSPTWHPSTPTVAQLCAQVEDTIKAVDRSCD----IGNF 236
 DB 588 FGADFMICENFPWLIEINSSPDGATTSVTARMCPQCLEDVVKVVIDRRTDPKALGNF 647
 QY 237 ELLWRQ 242
 DB 648 ELAYRQ 653
 RESULT 15
 ID O7PMD3 ANOGA PRELIMINARY; PRT; 501 AA.
 AC O7PMD3;
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DE ENGANGP0000022334 (Fragment).
 GN ORENames=ENSANGG00000019845;
 OS Anopheles gambiae str. PEST.
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicidae;
 OC Anophelinae; Anopheles.

```
OX NCBI_TaxID=180454;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RT "Anopheles gambiae re-annotation.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.
CC EMBL; AAAB01008980; EAA13971.3; -; Genomic DNA.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR011761; ATP_GRASP.
DR InterPro; IPR004344; Tub_cyr_lygase.
DR Pfam; PF03133; TTL; 1.
DR PROSITE; PSS0975; ATP_GRASP; 1.
FT NON_TER 1
FT NON_TER 501
SQ SEQUENCE 501 AA; 58862 MW; 806539A2FD8E76DF CRC64;

Query Match 40.1%; Score 521; DB 2; Length 501;
Best Local Similarity 45.9%; Pred. No. 4.4e-38;
Matches 112; Conservative 41; Mismatches 81; Indels 10; Gaps 4;

QY 1 IDGIRNIWIKPAKSRGRDIVCMDRVETLELAADHPLSRDNKXVVKYIETPLLICD 60
DB :|||:||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
262 IDGIRNMWIKPGNCRGGLGIMLFNDNRKLEHVD-----NPDVYVAQKYIERPLLIHC 317
QY 61 TKFDIROWFLVT-DWNPFTTWFKESYLRFSTQRFSLDKLDSAIHLCNNAVQK-YLKNDV 118
DB :|||:||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
318 TKFDIROYFLITTYNNVLKWMYRNCYLRFSSRQFNLDLDFSEIHLTNYSIQKNYAKEVR 377
QY 119 GRSPLLPAHNMWTSTRQEYLQGRGAVWGVSVIYPSMKKAIAMKVAQDHVEPRKNSF 178
DB :|||:||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
378 EGADALPASNNWSLKRFOEHLQSLDKGFYWERKIYPMKKNIILAVCASLDGKMERNMF 437
QY 179 ELYGADFVLGRDRPWLIEINSSPTMHPSTPVTQAQLCAQVOEDTIKVADRESCD----IG 234
DB :|||:||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
438 ELYGADFMVTNFTMLIEINTSPDLSSSTDVT7SVICPAVLEDLVKVVIDNTKDKRAGTG 497
QY 235 NFEL 238
DB :|||
498 QFEL 501
```

Search completed: April 4, 2006, 12:45:57
Job time : 36.6599 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:45:31 ; Search time 8.99628 Seconds
(without alignments)
2588.231 Million cell updates/sec

Title: US-10-635-977-2_COPY_133_374
Perfect score: 1299
Sequence: 1 IDGLRNIIWIKPAKSRGRD.....IKVAVDRSCDIGNFELLWRQ 242

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80: *
1: Pir1: *
2: Pir2: *
3: Pir3: *
4: Pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	850	65.4	352	2 T12515	hypothetical prote
2	247.5	19.1	379	2 A45443	tubulin-tyrosine 1
3	243	18.7	640	2 E88575	protein ZK1128.6
4	243	18.7	680	2 T27699	hypothetical prote
5	209.5	16.1	1198	2 T20262	hypothetical prote
6	209.5	16.1	1203	2 C89217	protein C55A6.2 [i
7	156	12.0	662	2 T20343	hypothetical prote
8	155.5	12.0	883	2 A96805	hypothetical prote
9	155	11.9	403	2 T37571	tubulin-tyrosine 1
10	144	11.1	753	2 S48261	hypothetical prote
11	87	6.7	439	2 T01270	hypothetical prote
12	85	6.5	1872	2 T30888	vitellogenin - Ath
13	84.5	6.5	522	2 T45824	hypothetical prote
14	83.5	6.4	374	2 E89920	conserved hypothet
15	83.5	6.4	930	2 A25923	progesterone recep
16	82.5	6.4	349	2 F91218	probable transport
17	82.5	6.4	349	2 H86064	probable transport
18	82.5	6.4	349	2 D65182	probable transport
19	82	6.3	306	2 A25698	probable protein k
20	82	6.3	1124	2 JH0588	calmodulin-binding
21	81.5	6.3	717	2 AC1419	DNA topoisomerase
22	81	6.2	270	2 B84813	probable RING zinc
23	81	6.2	524	2 T3050	cyclin E - Caenorh
24	81	6.2	570	2 T30156	hypothetical prote
25	80	6.2	1577	2 T30858	glucosyltransferas
26	80	6.2	2176	2 T39188	probable U5 snRNP-
27	79.5	6.1	380	2 H81152	anticodon nuclease
28	79.5	6.1	433	2 T04594	aldehyde dehydroge
29	79.5	6.1	1707	2 S77908	hypothetical prote

30	79.5	6.1	2142	2 D86303	P17F16.1 protein -
31	79	6.1	733	2 A64650	flagellar biosynth
32	79	6.1	733	2 E71937	flagellar biosynth
33	79	6.1	742	2 F84643	hypothetical prote
34	79	6.1	1528	2 S13743	DNA strand transfe
35	78.5	6.0	286	2 G69537	ribosomal protein
36	78.5	6.0	478	2 T03750	violaxanthin de-ep
37	78.5	6.0	515	2 C71158	probable thermosta
38	78.5	6.0	857	2 T04208	probable anthranil
39	78.5	6.0	1139	2 T33368	hypothetical prote
40	78	6.0	1012	2 S68259	DNA polymerase gam
41	77.5	6.0	627	2 AE2714	hypothetical prote
42	77.5	6.0	637	2 B97496	hypothetical prote
43	77.5	6.0	751	1 TVVP7H	large T antigen -
44	77.5	6.0	950	2 A82986	adenylate cyclase
45	77	5.9	322	2 AF2725	transcription regu

ALIGNMENTS

RESULT 1

T12515
hypothetical protein DKFZp434B103.1 - human
C:Species: Homo sapiens (man)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 09-Jul-2004
C:Accession: T12515
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z17524
A:Accession: T12515
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-352 <WAM>
A:Cross-references: UNIPROT:O9Y4R7; UNIPARC:UPI00001377A6; EMBL:AL096725
A:Experimental source: adult testis; clone DKFZp434B103
C:Genetics:
A:Note: DKFZp434B103.1

Query Match	65.4%;	Score 850;	DB 2;	Length 352;	
Best Local Similarity	61.8%;	Pred. No. 7.1e-71;			
Matches 152;	Conservative 41;	Mismatches 49;	Indels 4;	Gaps 1;	
Qy	1	IDGLRNIIWIKPAKSRGRDIVCM	DRVEEILELAADHPLSRDNKVVQK	YIETPLLICD 60	
Db	61	MEGDRNIWIKPGAKSRGIMCDH	LEMLKLVNPNVWKGKVVQK	YIERPLIFG 120	
Qy	61	TKFDIRQWFLVDNPLTIWFKES	YLRFSTQRFSLDKLDSAIHL	CNNAVQKYLKNDVGR 120	
Db	121	TKFDIRQWFLVDNPLTIWVFRD	SYIRFSTQPSFLKNDNSVHL	CNNSIQKHLENSCHR 180	
Qy	121	SPLLPAHNMWTSRFOEYLQ	RQGVGSVIYPSMKKAI	AHAMKVAODHVEPRKNSFEL 180	
Db	181	HPLLPDNNWSQRFQALQEM	GAPNAWSTVIIPGMDAVI	HALQTSQDTVQCRKASFEL 240	
Qy	181	YGADFVLGRDRFRPWLIEINS	PTMHPSTFVTAQCAQVQ	EDTIKVAV ----DRSCDIGNF 236	
Db	241	YGADFVGEDFQPMLEINAS	PTMAPSTAVTARLCAGVQ	ADTLRVVIDRMLDRNC	DGTGAF 300
Qy	237	ELLWRQ	242		
Db	301	ELIYKQ	306		

RESULT 2

A45443
tubulin-tyrosine ligase (EC 6.3.2.25) [validated] - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A45443
R:Rersfeld, K.; Wehland, J.; Plessmann, U.; Dodemont, H.; Gerke, V.; Weber, K.
J. Cell Biol. 120, 725-732, 1993
A:Title: Characterization of the tubulin-tyrosine ligase.

A;Reference number: A45443; MUID:93147125; PMID:8093886

A;Accession: A45443

A;Status: preliminary

A;Molecule type: mRNA; protein

A;Residues: 1-379 <ERS>

A;Cross-references: UNIPROT:P38160; UNIPARC:UPI00001377A9; GB:X68453; GB:S54050; NID:521

A;Experimental source: brain

A;Note: sequence extracted from NCBI backbone (NCBIN:123858, NCBIP:123859)

C;Keywords: ligase

Query Match 19.1%; Score 247.5; DB 2; Length 379;

Best Local Similarity 29.4%; Pred. No. 4.4e-15; Indels 33; Gaps 9; Matches 69; Conservative 40; Mismatches 93;

QY 2 DGLRNIIKPAKSRGRDIVCMRDVEIELEAAADHPLSRDNK---WVQKYIETPLLI 58

DB 141 DGEVNVIAKSSAGAKGEGILISSEATELLDFI-----DNQGVHVIQYLERPLL 192

QY 59 --CDTKFDIRQWFLVTDWNPITTFYKESYLRPSTORFSLDKL-DSAIHLCNNAVOKYILK 115

DB 193 EPGHRKFDIRSVLVD--HQYNIYLYREGVLRITASEPYHTDNFQDKTCHLTNHCIOKEYS 250

QY 116 NDVGRSPLPAHNMWTSRFEYLQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPR- 174

DB 251 KNYGK---YEEGNEMFEEFNQVLT-----SALNITLESILLOIKHIIIRSCLLSVEPAI 302

QY 175 -----KNSPELYGADFVLGRDPRPWLIEINSSPTMHPSTPTVAQLCAQVQEDTI 223

DB 303 STRHLVPQSOLFQGFDFWDEDLKWLIEVNGAPAC--AQKLYAELCQGIVDIAI 355

RESULT 3

E88575

protein ZK1128.6 [imported] - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004

C;Accession: E88575

R;anonymous, The C. elegans Sequencing Consortium.

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog

A;Reference number: A75000; MUID:99069613; PMID:9851916

A;Note: see websites genome.wustli.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_ele

A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A;Accession: E88575

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-640 <STO>

A;Cross-references: UNIPROT:Q09647; UNIPARC:UPI0000017A077; GB:chr_III; PIDN:CAA87425.1;

C;Genetics:

A;Gene: ZK1128.6

A;Map position: 3

Query Match 18.7%; Score 243; DB 2; Length 640;

Best Local Similarity 27.8%; Pred. No. 2.3e-14; Indels 38; Gaps 9; Matches 67; Conservative 47; Mismatches 89;

QY 2 DGLRNIIKPAKSRGRDIVCMRDVEIELEAAADHPLSRDNKWWVQKYIETPLLI 61

DB 285 DASHRV-IVKPPASARGTGISVTRPKDPFTTATL-----VAQHVIERPLTNRA 333

QY 62 KFDIRQWFLVTDWNPITTFYKESYLRPSTORFSLDK---LDSAIHLCNNAVOKYILKND- 117

DB 334 KFDRLRYAVVPTFEPLRVYIDQGLVRFAVPYSHSVSTISNKMHLTNYINKLAADG 393

QY 118 VGRSPL--LPAHNMWTSRFEYLQGRGAVGWSVIYPSMKKAIAHAMKVAQD 169

DB 394 VANKPVPKWTLLHLW--EHFDEMGVDREKIQRE-----IEEVIKAFISTEKPIRE 442

QY 170 H----VPRKNSPELYGADFVLGRDPRPWLIEINSSPTMHPSTPTVAQLCAQVQEDTI 225

DB 443 HMSRFLQEFCYELFGIILDEDYKFWLLEVNISFLHSGLTFLDVSVKAPLAKDVLNL 502

QY 226 A 226

DB 503 A 503

RESULT 4

T27699

hypothetical protein ZK1128.6 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C;Accession: T27699

R;Berks, M.

submitted to the EMBL Data Library, January 1995

A;Reference number: Z20407

A;Accession: T27699

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-680 <WIL>

A;Cross-references: UNIPROT:Q09647; UNIPARC:UPI000017A076; EMBL:Z47357; PIDN:CAA87425.2

A;Experimental source: clone ZK1128

C;Genetics:

A;Gene: CBSP:ZK1128.6

A;Map position: 3

A;Introns: 114/1; 208/3; 242/3; 334/3; 362/2; 405/2; 424/2; 550/2; 633/2

Query Match 18.7%; Score 243; DB 2; Length 680;

Best Local Similarity 27.8%; Pred. No. 2.5e-14; Indels 38; Gaps 9; Matches 67; Conservative 47; Mismatches 89;

QY 2 DGLRNIIKPAKSRGRDIVCMRDVEIELEAAADHPLSRDNKWWVQKYIETPLLI 61

DB 325 DASHRV-IVKPPASARGTGISVTRPKDPFTTATL-----VAQHVIERPLTNRA 373

QY 62 KFDIRQWFLVTDWNPITTFYKESYLRPSTORFSLDK---LDSAIHLCNNAVOKYILKND- 117

DB 374 KFDRLRYAVVPTFEPLRVYIDQGLVRFAVPYSHSVSTISNKMHLTNYINKLAADG 433

QY 118 VGRSPL--LPAHNMWTSRFEYLQGRGAVGWSVIYPSMKKAIAHAMKVAQD 169

DB 434 VANKPVPKWTLLHLW--EHFDEMGVDREKIQRE-----IEEVIKAFISTEKPIRE 482

QY 170 H----VPRKNSPELYGADFVLGRDPRPWLIEINSSPTMHPSTPTVAQLCAQVQEDTI 225

DB 483 HMSRFLQEFCYELFGIILDEDYKFWLLEVNISFLHSGLTFLDVSVKAPLAKDVLNL 542

QY 226 A 226

DB 543 A 543

RESULT 5

T20262

hypothetical protein C55A6.2 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C;Accession: T20262

R;Kershaw, J.

submitted to the EMBL Data Library, October 1996

A;Reference number: Z19243

A;Accession: T20262

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1198 <WIL>

A;Cross-references: UNIPROT:O17720; UNIPARC:UPI000017A066; EMBL:Z81051; PIDN:CAB02862.2

A;Experimental source: clone C55A6

C;Genetics:

A;Gene: CBSP:C55A6.2

A;Map position: 5

A;Introns: 32/3; 68/1; 107/3; 201/3; 273/1; 364/3; 407/3; 460/3; 476/2; 534/2; 677/3; 7

Query Match 16.1%; Score 209.5; DB 2; Length 1198;

Best Local Similarity 25.5%; Pred. No. 6.4e-11; Indels 31; Gaps 9;

Matches 60; Conservative 57; Mismatches 87;

QY 8 WIIKPAKSRGRDIVCMRDRVEIILEAAADHPLSRDNKVVQKYYIETPLLCIDTKFDIRQ 67
Db 715 FIVKPTNSRQKGIFFANSMDI-----PAEGPL-----LVSRYLKDPYLVNNHKFDLRI 764
QY 68 WFLVTDWNPITWFKESYLRFSTORF--SLDKLDS--AIHLCNNAVOK-----YLKNDVG 119
Db 765 YVAVTSFPLVAVYVSEGLARLASRPYDTSASSADSNSEYVHLTNYSINKNSTSFVRNEM 824
QY 120 RSPLLPAHNMWTSRFOEYLQGRGAVGWSVIYPSMKKAIHAHMKVAQDHV--EPRKN- 176
Db 825 SSEDLL--GHKWTGLGALLRYVNEGKDA---KLMLRIEDLIVKSLLSIQNSVATASRTNL 879
QY 177 -----SPELYGADFVLGRDPRFWLIEINSPTMHPSTPTVTAQLCAQVQEDTIKVA 226
Db 880 RFACTNFELFGDVLVDQALKPWLLEVNLSPLACDAPLDSLLKTRLIADLNL 934

RESULT 6

C9217
protein C55A6.2 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: C89217
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; PMID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: C89217
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1203 <STO>
A:Cross-references: UNIPROT:O17720; UNIPARC:UPI000017A067; GB:chr_V; PIDN:CAB02862.1; PI
C:Genetics:
A:Gene: C55A6.2
A:Map position: 5

Query Match 16.1%; Score 209.5; DB 2; Length 1203;
Best Local Similarity 25.5%; Pred. No. 6.5e-11;
Matches 60; Conservative 57; Mismatches 87; Indels 31; Gaps 9;
QY 8 WIIKPAKSRGRDIVCMRDRVEIILEAAADHPLSRDNKVVQKYYIETPLLCIDTKFDIRQ 67
Db 720 FIVKPTNSRQKGIFFANSMDI-----PAEGPL-----LVSRYLKDPYLVNNHKFDLRI 769
QY 68 WFLVTDWNPITWFKESYLRFSTORF--SLDKLDS--AIHLCNNAVOK-----YLKNDVG 119
Db 770 YVAVTSFPLVAVYVSEGLARLASRPYDTSASSADSNSEYVHLTNYSINKNSTSFVRNEM 829
QY 120 RSPLLPAHNMWTSRFOEYLQGRGAVGWSVIYPSMKKAIHAHMKVAQDHV--EPRKN- 176
Db 830 SSEDLL--GHKWTGLGALLRYVNEGKDA---KLMLRIEDLIVKSLLSIQNSVATASRTNL 884
QY 177 -----SPELYGADFVLGRDPRFWLIEINSPTMHPSTPTVTAQLCAQVQEDTIKVA 226
Db 885 RFACTNFELFGDVLVDQALKPWLLEVNLSPLACDAPLDSLLKTRLIADLNL 939

RESULT 7

T20343
hypothetical protein D2013.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T20343; T22085
R:Mortimore, B.
submitted to the EMBL Data Library, January 1995
A:Reference number: Z19259
A:Accession: T20343
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-662 <WIL>
A:Cross-references: UNIPROT:Q09512; UNIPARC:UPI000013B847; EMBL:Z47808; PIDN:CAA87778.1;

A:Experimental source: clone D2013
R:Matthews, P.
submitted to the EMBL Data Library, January 1995
A:Reference number: Z19510
A:Accession: T22085
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-662 <WIL>
A:Cross-references: UNIPARC:UPI000013B847; EMBL:Z47809; PIDN:CAA87783.1; GSPDB:GN00020
C:Genetics:
A:Gene: CESP:D2013.9
A:Map position: 2
A:Introns: 78/2; 114/3; 330/3; 471/1; 521/2; 587/2

Query Match 12.0%; Score 155.5; DB 2; Length 662;
Best Local Similarity 25.8%; Pred. No. 2.7e-06;
Matches 62; Conservative 36; Mismatches 100; Indels 42; Gaps 11;
QY 1 IDGLRNIWIIKPAKSRGRDIVCMRDRVEIILEAAADHPLSRDNKVVQKYYIETPLLI-- 58
Db 431 LNGQHNWVIVKPNLARGMDMTVTEDLNQIIRMI-----ETGPKIVCEYIPRLLRPR 483
QY 59 ---CDTKFDIRQWFLVTDWNPITWFKESYLRFSTORFSLDKL--DSAIHLCNNAVQKY 113
Db 484 PDNGNKVKFDLRYIVFLNGIAPVTAYVYNRFWRFAINEFSLSNFEDVETHF---TVFNY 540
QY 114 LKNDVGRSPLPAHNMWTSRFOEYLQGRGAVGWSV---IYPSMKKAI-AHAMKVAQD 169
Db 541 L--DKEILQWKCN-----FIETIKAYPRIOSEVQKDNILTRIKAEIAAAKEEAPR 592
QY 170 HVEPRKNSFELYGADFVLGRD---FREWLIENSPTMHPSTPTVTAQLCAQVQ--DTI 223
Db 593 GVAPNVQSRAMVGVDMQLHGDNVKSITLLEINFMPD-----TTRACQYYPDFADTV 645

RESULT 8

A96805
hypothetical protein TSM16.14 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: A96805
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A96805
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-883 <STO>
A:Cross-references: UNIPROT:Q9CAQ1; UNIPARC:UPI00000A09A3; GB:AE005173; NID:g6382502;
C:Genetics:
A:Gene: TSM16.14
A:Map position: 1

Query Match 12.0%; Score 155.5; DB 2; Length 883;
Best Local Similarity 25.3%; Pred. No. 4.3e-06;
Matches 55; Conservative 36; Mismatches 93; Indels 33; Gaps 7;
QY 2 DGLRNIWIIKPAKSRGRDIVCMRDRVEIILEAAADHPLSRDNKVVQKYYIETPLLICDT 61
Db 641 DQLNNLWIKPNWARTDTSITDNLGAIIR-----MMETGPKICQKYIEHPALPKGN 693
QY 62 KFDIRQWFLVTDWNPITWFKESYLRFSTORFSLDKLDSAIHLCNNAVQKYLKNDVGRS 121
Db 694 KFDLRYVVLVRSIDPLFIYLIETFEVRLSNPNPYSLKHSFFFEYETHFTWNY-----GRK 748

QY 122 PLLPAHNMWTSTRFOYLQKQ-----RGAVGWSV-----IYPSMK-----KAIHAAMK 165
Db 749 -----LNHRKPAEVRFEQEHNDYAFHFVNNTIQLSIVKWMIDHEKXQVIRAVPEAAA 804
QY 166 VAQDHVEPRKNSPELYGADFLVGRDRFRPMLIEINSSP 202
Db 805 LAHPMQSPK-SRAMYGVDMLDSSPEPKLEVITYCP 840

RESULT 9
T37571
tubulin-tyrosine ligase related protein - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C/Accession: T37571
R;/Badcock, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, April 1996
A/Reference number: Z21726
A/Accession: T37571
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-403 <BAD>
A/Cross-references: UNIPROT:Q10438; UNIPARC:UPI000013A94B; EMBL:X70721; PIDN:CAA94694.1;
A/Experimental source: strain 972h-; cosmid c12B10
C/Genetics:
A/Gene: SPDB:SPAC12B10.04
A/Map position: 1
A/Introns: 320/3; 348/3

Query Match 11.9%; Score 155; DB 2; Length 403;
Best Local Similarity 21.3%; Pred. No. 1.7e-06;
Matches 54; Conservative 53; Mismatches 90; Indels 56; Gaps 10;

QY 8 WIIPKAASGRDIVCMRVEIELEL-----AAADHPILSRDNK----- 45
Db 144 YILKPSMCDRAQIGRLPFTTIEELQAIFDSFDDSESEAEGLBEKGDITVAFNKIVISQ 203
QY 46 ---WVQKYIETPLLICDTKFDIRQWFLVTDNMPLTIWFKYESYLRFSTQRFSLDKLDSA 102
Db 204 IRNFLVQKYISKPLLDHRRFHIRAYVLAT--GALSVLFENMLCLLARDKYKKTPDPD 261
QY 103 I---HLCNNAVQKYLKNDVGRSPLLPANHNMWTSTRFOYLQKQRGAVGWSVLYPSMKKA 159
Db 262 LLFSHLSNTCLQ---GDNVQSST---RDFWNTS-----IENK-----DDIFKSLNT 303
QY 160 IAHAMKVAQD---HVEPRKNSPELYGADFLVGRDRFRPMLIEINSSPTMHPSTPVTQAQLC 215
Db 304 IGDVFEAAATQGIHFQPLENCFEIFGVDFLVDCESQVYLLEVNS---YPDFKQTKGNL 359
QY 216 AQVOEDTIKVAVD 228
Db 360 SNIENLFSVAVE 372

RESULT 10
S48261
hypothetical protein YBR094w - yeast (Saccharomyces cerevisiae)
N/Alternate names: hypothetical protein YBR0821
C/Species: Saccharomyces cerevisiae
C/Date: 01-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 09-Jul-2004
C/Accession: S48261; S45962; S41800; S44676
R;/Mannhaupt, G.; Stucka, R.; Ehnlé, S.; Vetter, I.; Feldmann, H.
Yeast 10, 1363-1381, 1994
A/Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
A/Reference number: S48255; MUID:95208357; PMID:7900426
A/Accession: S48261
A/Status: nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-753 <MAN>
A/Cross-references: UNIPROT:P38254; UNIPARC:UPI000013A3DD; EMBL:X78993; NID:g476045; PID
R;/Feldmann, H.; Mannhaupt, G.; Schwarzlöse, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994

A/Reference number: S45927
A/Accession: S45962
A/Molecule type: DNA
A/Residues: 1-753 <PE2>
A/Cross-references: UNIPARC:UPI000013A3DD; EMBL:X35963; NID:g536366; PIDN:CAA85047.1; P
R;/Dekker, P.J.T.; Hoekert, W.; van Oosterum, K.; Grivelli, L.A.
submitted to the EMBL Data Library, December 1992
A/Reference number: S31313
A/Accession: S41800
A/Molecule type: DNA
A/Residues: 167-351, 'TPE', '356-449, 'A', '451-562, 'R', '564-753 <DEK>
A/Cross-references: UNIPARC:UPI0000168D37; EMBL:X69881; NID:g4086; PIDN:CAA49508.1; PID
C/Genetics:
A/Cross-references: SGD:S0000298
A/Map position: 2R

Query Match 11.1%; Score 144; DB 2; Length 753;
Best Local Similarity 21.8%; Pred. No. 4.1e-05;
Matches 57; Conservative 44; Mismatches 105; Indels 56; Gaps 9;

QY 8 WIIPKAASGRDIVCMRVEIE-----LELAAADHP----- 39
Db 476 WIVKPSMSDKGQIGRVFKTIEDLQAIFDSFDDSEAEESGNDDDADDVNGEPMDNKNVN 535
QY 40 LGRDNKWVYQKYIETPLLIC---DTKFDIRQWFLVTDNMPLTIWFKYESYLRFSTQRF-- 94
Db 536 ISQLRHFIIQEYLTNPPLLLASMDNKKFIRCY--VVCRGDLQVVFVDRMLALPAAKPFVP 593
QY 95 -----SLDKLDSAHLCCNNAVQKYLKNDVGRSPLLPANHNMWTSTRFO--EYLQKQGRG 145
Db 594 LDPYAYSVTDLKDLECHLTNTCLQS--KKDKDQSSVL-----EFDISIEIPNERKS 642
QY 146 AWGSVVIYPSMKKATAHAMKVAQDHVEPRKNSPELYGADFLVGRDRFRPMLIEINSSPTMH 205
Db 643 NI-KEQIHSITNDVFLAAVNAVRLNFQPLNPAFETYGVDFLIDSNYEYVKLLEINAFPDFK 701
QY 206 PSTFVTVAQLCAQVOEDTIKVAV 227
Db 702 QTGKDLKNLIDELFDDTVKYCV 723

RESULT 11
T01270
hypothetical protein At2g19220 [imported] - Arabidopsis thaliana
N/Alternate names: hypothetical protein F27F23.2; hypothetical protein T20K24.23
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C/Accession: T01270; T00543; A84574
R;/Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kau
submitted to the EMBL Data Library, May 1998
A/Description: Arabidopsis thaliana chromosome II BAC F27F23 genomic sequence.
A/Reference number: Z14177
A/Accession: T01270
A/Status: translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-439 <ROU>
A/Cross-references: UNIPROT:O65915; UNIPARC:UPI00000A9775; EMBL:AC003058; NID:g3135250;
R;/Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
submitted to the EMBL Data Library, July 1997
A/Description: Arabidopsis thaliana chromosome II BAC T20K24 genomic sequence.
A/Reference number: Z14167
A/Accession: T00543
A/Status: translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-439 <ROW>
A/Cross-references: UNIPARC:UPI00000A9775; EMBL:AC002392; NID:g3176701; PID:g3176720
R;/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
Muss, D.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84574
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-439 <STO>
A:Cross-references: UNIPARC:UPI00000A9775; GB:AE002093; NID:g3135252; PIDN:AAC16452.1; G
C:Genetics:
A:Gene: At2g19220; F27F23.2; T20K24.23
A:Map position: 2
A:Introns: 94/3; 263/3

Query Match 6.7%; Score 87; DB 2; Length 439;
Best Local Similarity 23.4%; Pred. No. 3.8;
Matches 57; Conservative 29; Mismatches 96; Indels 62; Gaps 14;
QY 22 VCMRVVEE-----ILEAAADHPLSRDNKVVQKYIETPLLCDFD-----IRQW 68
DB 136 ICEHDEENVGVPLRSASDIATRRHYKWSSSH-----AIVDTFCQESLKGIRPKN 191
QY 69 FLVT--DW-----NPLTIWFKSYL--RFSTQRFSLDKLDSAIHLCLNNAVQKYLKND 117
DB 192 HLFTKESKWKILEKINRITGLGYTHKQLENHFTRTTSWK-----HWCETIASPIMKWD 245
QY 118 VGRSPLLPAAHNMWTSRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVA-----QDHVPR 174
DB 246 ANTRKFGATEDW-----DKYLMINKRARVF-----KRRHIPHADKLATIFKGRIBPG 293
QY 175 KNSFELYGADFVLGRDFRFLWEINS-SPTMHPSTPTVAQLCAQVQEDTIKAVDRSCDI 233
DB 294 KTK-----TRYKRVRVIDHSESPLQHDHQTPTSSVVVNTNE-PVKGSDDRASED 341
QY 234 GNFE 237
DB 342 GNVE 345

RESULT 12
T30888
vitellogenin - Athalia rosae
C:Species: Athalia rosae
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 09-Jul-2004
A:Accession: T30888
R:Nose, J.; Lee, J.; Ueno, T.; Hatakeyama, M.; Oishi, K.
Insect Biochem. Mol. Biol. 27, 1047-1056, 1997
A:Title: Cloning of cDNA for vitellogenin of the parasitoid wasp, Pimpla nipponica (Hym
A:Reference number: Z20922; MUID:98231104; PMID:9569645
A:Accession: T30888
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1872 <NOS>
A:Cross-references: UNIPROT:Q17083; UNIPARC:UPI0000083E90; EMBL:AB007850; NID:d1170514;
C:Genetics:
A:Gene: Vg
C:Superfamily: boll weevil vitellogenin

Query Match 6.5%; Score 85; DB 2; Length 1872;
Best Local Similarity 21.1%; Pred. No. 38;
Matches 57; Conservative 33; Mismatches 118; Indels 62; Gaps 10;
QY 19 RDIVCMRVVEILELAADHPLSRDNKVV-QKYIETPLLC---DTKFDIOWF--LV 71
DB 1113 KDILSEKVPQKGYALISERGSEQNEWAFQKSTGLALRFTSEGEQTEHQLRNWAMEL 1172
QY 72 TDWNPITWFKSYLRFSTQRFSLDKLDSAIHLCLNNAV-----OK 112
DB 1173 QSEDLAFAFLDDEIQHSNNLWLD-----GQHSNEAVLVTASVANYTKMGGSDSQS 1228
QY 113 YLKNDVGR-----SPLPA-----HNMWTSRFOEYLQROGRGAV-----W 148
DB 1229 GYQHPAGRKNSGGEDNQSSFTVFAPQPSQKPPDSESQELLOKAASGIKDAYSQSIDF 1288
QY 149 GSVIYPSMKKAIAHAMKVAQDHVPRKNSFELYGADFVLGRDFRFLWEINSPTMHPST 208

Db 1289 GAVFOGKKAEBYAATAFAMANSVPDPKSRVVFYGSNPAQOKNYQICL-----SAQSSRPKV 1344
QY 209 PYTAOLCAQVQEDTIKVAVD-----RSCDIG 234
DB 1345 PVWDFIKALKADPTTKFAADLKFGEQCSG 1374
RESULT 13
T45824
hypothetical protein F2K15.50 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
A:Accession: T45824
R:Rieger, M.; Gabel, C.; Mueller-Auer, S.; Schaefer, M.; Zipp, M.; Mewes, H.W.; Lemcke
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23015
A:Accession: T45824
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-522 <RIE>
A:Cross-references: UNIPROT:Q9M3B3; UNIPARC:UPI000000C655; EMBL:AL132956
A:Experimental source: cultivar Columbia; BAC clone F2K15
C:Genetics:
A:Map position: 3
A:Introns: 69/3; 271/3; 294/1; 327/3; 397/3; 443/3
A:Note: F2K15.50
C:Superfamily: Arabidopsis thaliana hypothetical protein F2K15.70

Query Match 6.5%; Score 84.5; DB 2; Length 522;
Best Local Similarity 21.8%; Pred. No. 8.1;
Matches 48; Conservative 31; Mismatches 62; Indels 79; Gaps 11;
QY 5 RNIMWIKP--AAKSRGRDIVCMRVVEILELAADHPLSRDNKVVQKYIETPLLCITDK 62
DB 199 RLLMLVKVITVAIILGLNTVC-DALD-----FIVTLFVKOTE 235
QY 63 FDIQWFLVTDWNPITWFKSYLRFSTQRFSLDKL-----DSAIHLCLNNAV 110
DB 236 TPIKGDFLST-----KSKQLRLVHRTVSLDDIKLTKNAMNMINDVVLGVTQAGL 285
QY 111 QKYLKNDVVG-----RSPLLPAHNMWTSRFOEY--LQROGRGAVGWS 150
DB 286 SYLARRYGEETKNRQKQKLPKRIRLSALLV--NLRPITGIIQDLADMMEKSGKCRWGN 343
QY 151 ----VIVP---SMKKAIAHAMKVAQDHVPRKNSFELYGA 183
DB 344 WFGVVPFPSTALRDDPLEHLEIAQKTISSRKNS---YGA 380

RESULT 14
E89920
conserved hypothetical protein SA1262 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
A:Accession: E89920
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Og
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: E89920
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-374 <KUR>
A:Cross-references: UNIPROT:Q99U59; UNIPARC:UPI00000D76B2; GB:BA000018; PID:g13701227;
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA1262

Query Match 6.4%; Score 83.5; DB 2; Length 374;
Best Local Similarity 26.4%; Pred. No. 6.5;
Matches 42; Conservative 19; Mismatches 59; Indels 39; Gaps 7;

```
QY      83 KESYLRFSTQRFSLDKLDSAIHLN-NAVOKYLNVDGRSPLLPANHMTSTRFOEYLOR 141
      :|: || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      35 EETQPRFINQLLSIDGITSIFVWNFLAVDKPKAD----- 70
      :|: || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      142 QGRGAVWGSVIYPSMKKATAHAKVAQDHVEPR-KNSFELYGADFVLGRDRFPWLJEINS 200
      :|: || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      71 -----W-EVILPDIKAAFSADANKVLESVNEPQIDNHFGIEIKAEILTFKGI-PYQIKLTS 122
      :|: || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      201 SP-----TMIPTSTPTAQLCAQVQEDTIKVAVDRSCDIGN 235
      :|: || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      123 ADQELREQLPQTYVDHMTQAQTAHDNI-VFMRKWLDLGN 160
      :|: || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
A25923
progesterone receptor - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 09-Jul-2004
C;Accession: A25923
R;Loosfelt, H.; Atger, M.; Misrahi, M.; Guiochon-Mantel, A.; Meriel, C.; Logeat, F.; Ben
Proc. Natl. Acad. Sci. U.S.A. 83, 9045-9049, 1986
A;Title: Cloning and sequence analysis of rabbit progesterone-receptor complementary DNA
A;Reference number: A25923; MUID:87067449; PMID:3538016
A;Accession: A25923
A;Molecule type: mRNA
A;Residues: 1-930 <LOO>
A;Cross-references: UNIPROT:P06186; UNIPARC:UPI00001321A6; GB:M14547; NID:g165631; PIDN:
C;Superfamily: progesterone receptor; erba transforming protein homology
C;Keywords: DNA binding; nucleus; steroid hormone receptor; transcription regulation; zi
F;566-826/Domain: erba transforming protein homology <ERBA>
F;568-588/Region: zinc finger
F;604-628/Region: zinc finger
```

```
Query Match          6.4%; Score 83.5; DB 2; Length 930;
Best Local Similarity 26.0%; Pred. No. 21;
Matches 33; Conservative 20; Mismatches 45; Indels 29; Gaps 5;

QY      88 RFSTQRFSLDKLDSAI-----HLCNNAVOKYLNVDGRSPLLPANHMTSTRFOEYLORQ 142
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      273 RFSAPRVSLAEQDAPVAPGRSPLATSVV-----DFTHVPILPLNHAFLATRTRQLLEGE 326
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      143 G--RGAVWGSVIYPSMKKATAHAKVA-----ODHVEPRKNSFELYGADFVLGRDF 191
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      327 SYDGAAGAAASPFPVQGRGPSASSTPVAGDPDCTYPDAPKDDAPPLYG-----DF 379
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      192 RPWLIEI 198
      :|: || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      380 QPPALKI 386
      :|: || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: April 4, 2006, 12:51:55
Job time : 10.9963 secs